

调查研究方法

SURVEY RESEARCH METHODS

（第3版 校订新译本）

弗洛伊德·J. 福勒(Floyd J.Flowe,Jr.) 著

孙振东 龙 藜 陈 荟 译

沈崇麟 校



万卷方法是重庆大学出版社自2004年7月以来陆续出版的一套深入、系统地介绍社会科学研究方法的书系，至今已出版图书50余个品种。

“万卷方法”的理想

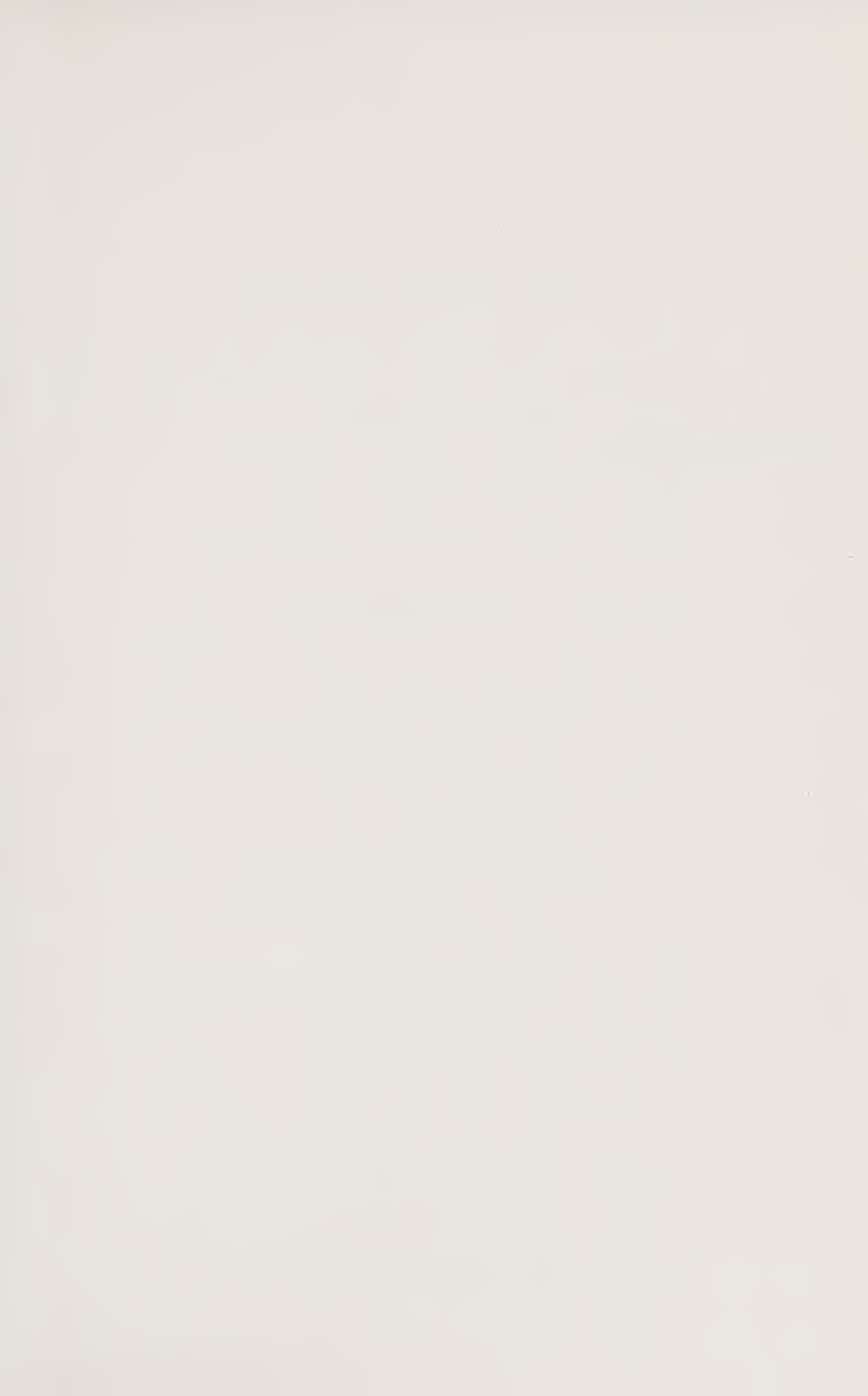
为社会学、政治学、管理学、经济学、传播学、心理学、教育学等社会科学各领域的学者和研究生提供一个内容规范、使用便捷的“研究方法工具箱”。

“万卷方法”的受众

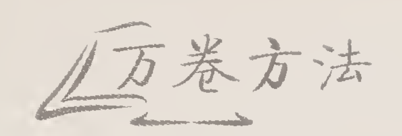
- 图书馆及大学社会科学各院系的资料室。
- 社会科学各领域的研究人员、研究生及高年级本科生。
- 社会及市场调查的实务工作者。

万卷方法是一个开放的体系，欢迎投稿！

万卷方法与学术规范博客圈:
http://q.blog.sina.com.cn/fafang



万卷方法学术委员会



调查研究方法

SURVEY RESEARCH METHODS

（第3版 校订新译本）

弗洛伊德·J.福勒 著

孙振东 龙 藜 陈 荟 译

沈崇麟 校

重庆大学出版社

图书在版编目(CIP)数据

调查研究方法/(美)福勒(Flower,F. J.)著;
孙振东,龙藜,陈荟译. —3 版(校订新译本). —重庆:重庆大学出版社,2009.4(2013.5 重印)
(万卷方法)
书名原文:Survey Rearch Methods
ISBN 978-7-5624-3289-0

Ⅰ.调… Ⅱ.①福…②孙…③龙…④陈… Ⅲ.调查研究—方法 Ⅳ.C31

中国版本图书馆 CIP 数据核字(2009)第 040507 号

调查研究方法
(第 3 版 校订新译本)
(美)弗洛伊德 · J. 福勒,Jr 著
孙振东 龙 藜 陈 荟 译
沈崇麟 校
策划编辑:雷少波 林佳木
责任编辑:雷少波 版式设计:雷少波
责任校对:文 鹏 责任印制:赵 晟
*
重庆大学出版社出版发行
出版人:邓晓益
社址:重庆市沙坪坝区大学城西路 21 号
邮编:401331
电话:(023) 88617183 88617185(中小学)
传真:(023) 88617186 88617166
网址:http://www. cqup. com. cn
邮箱:fxk@ cqup. com. cn (营销中心)
全国新华书店经销
重庆华林天美印务有限公司印刷
*
开本:940 × 1360 1/32 印张:6.75 字数:174 千
2009 年 4 月第 2 版 2013 年 5 月第 4 次印刷
印数:9 001—10 500
ISBN 978-7-5624-3289-0 定价:25.00 元

作者简介

弗洛伊德·J.福勒,Jr 毕业于 Wesleyan 大学,于 1966 年获得密歇根大学社会心理学博士学位。此后,他大多数时间都是在马萨诸塞—波士顿大学的调查研究中心度过的。他在这个中心参加了许多调查项目的设计和施行,这些项目的主题极其广泛,包括:对地方政府和机构的态度、赌博、种族冲突、犯罪的恐惧、犹太人的识别、冒险的理解以及从健康计划中获得医疗服务的经历等。除此之外,还涉及其他更多的研究主题。最近他的方法研究主要集中在以下三个领域:调查问题的设计和评估、如何减少调查资料里与访谈员有关的误差和如何测量医疗服务对病人的影响。他在哈佛的公共健康学院以及其他地方教授调查研究方法。他担任调查研究中心的主任,长达 14 年之久。除了已发表的大量研究论文外,他还是《标准化调查访谈》(*Standardized Survey Interviewing*,与 T. W. Mangione 合作)和《改进调查问题》(*Improving Survey Questions*)两本被广泛使用的有关调查研究方法的专著的作者。

英文版前言

本书这个版本的出版目的与其前两个版本一样，都是为那些原本不是统计学者或方法论学者的人们提供一个有关调查中误差来源的一般知识的全面总结。我认为，那些重要的知识，诸如良好的样本、高回答率、好的问题设计、良好的访谈过程、有效的资料收集方案，以及它们对调查资料质量的影响等，从本书的第 1 版发行至今 16 年来并没有太多变化。弄清这些问题的重要性以及它们如何影响了调查估计质量，对于开展调查或者使用他人收集的调查资料而言都是非常重要的。对此，相信读者在读完本书后就会明白。

我做了相当大的努力以使本书对一般的读者来说都能清晰易懂。对社会科学研究的了解会帮助读者理解本书，但并非一定要有相关知识的人才能读懂本书。

本书是简写本，必然地，我就要对文章内容的详细程度做出选择。如果有读者希望对特定的问题诸如抽样、调查手段设计或者资料收集方案等有更深入的了解，那就需要进一步参看一些大部头的专著。

这一版本与前面版本相比，主要有三方面的不同：第一，本书包括了在最近 8 年里发表的方法论资料。第二，计算机和互联网的发展在给研究者在调查方法上提供了更多选择的同时，也使调查的具体操作有了很大变化。这些内容的加入则使本书更有新意。第三，调查问题的评估技术也在不断地进步之中。这是调查方法中一个处在持续不断发展中的领域。而本书的修订版则反映了这些进步。

在以前的版本里，我对罗伯特·凯恩(Robert Kahn)、莫里斯·阿克塞尔罗德(Morris Axelrod)以及后来的安格斯·坎贝尔(Angus Campbell)所给予的帮助表达了我诚挚的谢意。新版本又多处参考了罗伯特·格雷夫斯(Robert Groves)、西摩·萨德曼(Seymour Sudman)、诺曼(Norman)、布拉德伯恩(Bradburn)和唐·迪尔曼(Don Dillman)的著作。读了本书之后，想更多地了解调查研究方法的学生应当再去阅读上述人们的专著，这些专著都对方法论知识做出了很大的贡献。最后，我成长为一个方法论学者深受与查尔斯·F. 坎内尔(Charles F. Cannell)共事的影响，他对我们理解访谈员和调查手段设计的贡献极其巨大，他的工作对我的调查研究方法观点的影响是深远的。

关于本书，我要感谢玛丽·艾伦·科尔顿(Mary Ellen Colton)、汤尼·罗曼(Tony Roman)、维基·斯特林费洛(Vickie Stringfellow)和加拉赫·加拉格尔(Trish Gallagher)对本书的仔细审阅并对很多章节做出的有益的评点；感谢调查研究中心为我完成本修订版提供的时间和援助；感谢朱蒂·钱布利斯(Judy Chambliss)一直以来给予我的帮助，她使我在写作过程中得以保持健康的精神状态。尽管人们给予我巨大、有益的帮助，但本书的最终责任应由我个人承担。

目录

第1章 导 论

本书是一本关于调查(survey)的规范方法和实际操作步骤的专著。本书所谓的“调查”特指那些旨在借助提问(通常是对样本中的被调查对象提问),并在此基础上进而对调查对象在统计上加以描述这样一种调查。这样的调查熔抽样方法、问题设计和访谈技术为一炉。那些有意于收集、分析,或阅读调查资料的人,都必须详细了解调查的每一方面究竟是如何对调查的精确性(precision)、准确性(accuracy)和可靠性(credibility)产生影响的。

本书是一本有关社会调查资料收集方法的专著。内容包括资料收集的一般步骤、优良的操作规范，以及获得高质量的调查数据的各种设计策略的运用等。本书的主要目的是：为那些旨在收集、分析和阅读调查资料的人对资料收集的步骤作评估提供一个可靠的依据。读者将会逐渐了解那样一些方法，这些方法的每一细节都与他们可以从数量上把握的置信度和以调查数据为根据的统计量有关。

可以称为调查的资料收集和量度的方法有许多。但本书的重点则在具有下面这样一些特点的调查：

- 调查的目的在于得到统计数字，即对研究总体某些方面的定量或数字的描述；
- 收集信息的主要方法是对被调查人提问题，而被调查人给出的答案则构成了我们要分析的资料；
- 一般讲，收集的信息只涉及总体的一部分，即所谓的样本，而非总体中的每一个成员。

调查原因

美国的宪法规定每隔十年必须进行一次满足确定标准的调查。在这十年一度的普查中，通过向人们提出各种问题，得出各种有关总体的统计数字。由于未作抽样，因此必须假设资料的收集包括总体中的每一个人。

十年一度的普查目的是清点人口，以为众议院席位的分配提供依据，确保众议院席位的代表性。然而，这样一种普查也逐渐成为满足其他目的的主要信息来源。因此在简单的清点人口之外，人们开始用它来收集有关种族、年龄、住户构成、教育、居住方式及其他许多有关被清点人口特征的资料。

十年一度的普查的内容虽然已经有了很大的扩充，以满足政府机构和研究人员对描述性资料的需要，但它还是只涉及了

人们想从总体中了解的事物的一小部分。更何况它十年才进行一次，因此它的价值就更加有限。为了弥补这些不足，自 20 世纪 30 年代以来，各种有特定目的的调查逐渐流行起来，并成为美国社会生活的一部分。

绝大多数人可能都熟悉调查技术的三种用途：报刊杂志上的公众舆论调查；在大选中为候选人提供帮助的公众舆论调查；旨在了解消费者的消费偏好和兴趣的市场调查。这些日臻完善的调查研究项目中的每一种，其目的都在洞察公众的主观感觉。此外，得到有关人们的行为和处境的众多事实的唯一方法，是向样本中的个体本人提问。在公共政策领域内，已鲜有调查研究尚未涉足的地方。以下是调查研究涉足的主要领域的一个简明的清单：

- 由劳工统计局定期公布的失业率，以及其他许多有关职业和工作的统计数字都是以普查局所作的调查为依据的。与此同时，某些商业和工业调查则为我们提供了有关产品和人力需求方面的统计数字。
- 收入和消费方式则构成了另一个只有调查才能得到可靠数据的领域。消费者的花费方式和期望已被证明是经济趋势的重要前测指标。
- 自 20 世纪 50 年代以来由普查局进行的，旨在了解公共卫生服务设施的国家健康卫生访谈调查。这一调查收集了有关卫生健康状况、医疗设施的使用和可能引发疾病的行为等方面的基本数据。这些题目无一不属于只有通过调查才能得到可靠资料的领域。
- 有关犯罪问题的资料的主要来源历来都是警察局的记录。然而，警察局的记录只包括那些已经报案的案件，对大多数涉及受害者的犯罪问题而言，调查为我们提供了更为可靠的犯罪率量度和犯罪方式的分布状况。发端于 70 年代的全美犯罪调查便为我们提供了诸如这样的数字。调查也是量度人们对犯罪的忧虑和恐惧的唯一途径。

- 调查方法最早的应用者之一是美国农业部。农业部对农民进行调查,估计各种不同种类的谷物的种植率,并预测各种不同食品的产量。

精神卫生、运输需求和运输方式、政治行为、住房问题(包括房价和住房的实用性)、工人的满意程度等领域也广泛使用调查。无疑,美国最大的调查资料的收集者是联邦政府,尤其是其中的普查局和农业部。此外,大学、非赢利和赢利的调查组织每年都要进行数以千计的调查。

资助一项特定目的调查是一种昂贵的解决信息需求问题的方法。正因为如此,在启动这样一个项目前,我们必须审慎考虑是否有可能从已存的记录或其他来源收集相同的信息。尽管我们把调查看作试图得到总体的有关信息的各种途径中的首选,但我们更主张一个整尺度的概率抽样调查必须在确定所需信息无法通过其他途径得到之后方可实施。即使我们采取一种如此保守的方法,在一般情况下,我们还是会发现,只有一个特定目的的调查才能为我们提供我们所需要的信息。一个实施得当的调查除了能使我们得到所需的、用其他方法无法得到的信息之外,用它得到的数据还有以下三个潜在的特性,这些特性使得用它得到的信息的质量优于从其他途径得到的信息:

- 概率抽样使我们得以确信样本是无偏的,并估计数据的精确度可能是多少。(来自经由适当程序抽取的样本的数据的质量,较之那些来自与会者、大声说话者、投书者或刚好便于调查者组成的样本的数据有很大改进)
- 对所有回答人采取的完全一致的标准化量度,确保来自被量度的每一个人的信息都具有可比性。没有这样一种量度,就不可能产生有意义的统计量。
- 也许特定目的的调查是确保满足某一特定的分析所需的全部数据完整有效和相互关联的唯一方法。有时,我们可能会有全体事件中某一组事件的信息,但它们未必与作某

一种特定分析要求的其他特点相匹配。例如,医院出院结账的记录,无一例外都不包含有关收入的资料。正因为如此,为了研究收入和住院经历的关系,我们就必须作一个特定的调查,同时收集有关收入和住院这两方面的资料。

对任何一个特定的题目而言,我们总是可以从人们的议论、印象或某些记录中得到一些资料,但这些资料总是不无缺憾的。除了对资料的需求作评估以外,调查决策也必须考虑一些其他的问题,如调查人员的来源问题。除非我们拥有足够的资源聘用调查所需的工作人员和专家,否则我们就无法保证调查能得到质量上乘的数据。而这些问题就是下面我们所要介绍的内容——怎样构建一个完美的调查?

调查的组成

与所有科学的量度一样,社会调查的量度也非无误差的。实施一次调查的实际程序对相似性,即调查得到的数据描述它们所要描述的对象的准确性会有很大影响。

一个样本调查通常由三个不同的方法论领域组成——抽样、问题设计和访谈。这三者中的每一种,在样本调查之外各自都有许多应用,但就一个完美的调查设计而言,最重要的是将这三者组合成一体。

抽 样

普查意味着要收集总体中每一个个体的资料。要使调查数据能有实际使用价值,最主要的一环是懂得如何抽样,即从总体中抽取一个代表总体的子集。一个完备的抽样的关键在于找到一个给总体中的所有(或几乎是所有)成员以相等(或是已知)的被抽取的机会,和使用概率方法来抽取样本的途径。早期的调查和民意测验常常依靠方便的样本,或从一个显然排

除了总体中的某一部分成员的清单来抽取样本。用诸如这样的抽样方法进行的调查肯定提供不了可靠和可信的数字。

农业部实际上已创造出一种用于从居住在确定地区的总体中抽取可对之作可靠的统计描述的全概率样本的方法。这种抽样方法是从一种原来为抽取一个土地样本,以预测作物的收成这样一种设计逐渐发展起来的。抽取居住单位和从居住单位内抽取人,无非是它的延伸。第二次世界大战期间。农业部一些社会科学家做了一些与战争有关的社会调查。从那时开始,区域概率抽样才逐渐在社会调查中有了不可动摇的地位,并被普遍运用于一般总体的抽样。区域抽样至今仍是在入户个人访谈调查中使用的基本方法。康弗斯(Converse,1987)为我们详细描绘了调查方法在美国的发展过程。

自 20 世纪 50 年代以来,抽样技术一直在不断地发展,其中最为重要的发展是用电话随机数码拨号法抽取住户的方法。总之,长期以来抽样原理和实践都已经有了长足的进步。

问题设计

用问题作为测量手段是调查方法的又一基本特点。最初的调查无非是报刊杂志采访的延伸,并不注意问题的设计。但不久问题便暴露无遗,因为派调查员出去,用一组没有文字和操作上统一定义的问题作访谈,得到的答案存在很大差异。有鉴于此,20 世纪初,有些研究者开始编写标准化问题来测量主观意识现象。后来,在 1940 年美国农业部的研究人员将标准化问题延伸至真实或客观的问题的探索上,他们的做法也得到人们的认同。1951 年,佩恩(Payne,1951)的一本具有划时代意义的书问世,该书为我们提供了如何在问卷中撰写条理清晰的、调查人员可据此照本宣科地进行实地访谈的问题的实用指南。一般认为,是李克特(Likert,1932)将心理物理学方面的心理学家开发的用以测量主观现象的复合量表技术(例如:Thurstone & Chave,1929)运用于实用社会调查研究的实践中。

在过去 20 年里,问题设计上的主要进步是改进了对问题

进行评估的策略。与以前相比,现在越来越多的研究者都会对问题进行测试,以确定它们是否易于理解,以及答案是否有意义。通过分析被调查者的录音磁带来找出有疑问的问题,调查的试调查已变得更为系统化。这样,使得问题的措辞造句的选择变得更加客观,大大减少了调查人员的主观随意性。

访 谈

虽然并非所有的调查都包括访谈(因为一些调查只需调查对象自己填写问卷),但是由访谈员提问并记录答案的方法仍是一种普遍使用的方法。当派访谈员进行访谈时,最重要的问题是既要避免他们影响被调查人对问题的回答,又要尽可能地增大被调查人回答问题的准确性。

要使访谈员始终保持一致性,首要方法是给他们标准化的问题,其次是要注意对访谈员的培训,让他们学会如何操作调查,避免在收回的答案中引入大量的偏倚(Friedman,1942)。1954 年,海曼(Hyman)、弗尔德曼(Feldman)和斯滕伯(Stember)出版了一系列研究报告,在这些报告中,他们列举了问题的文字编写之外的那些可能导致访员会对得到的答案产生影响的种种因素。他们的报告促使我们在访员培训时,注意如何使访员在保证使答案没有偏倚的前提下,在遇到答案不够完整时使答案完整,以及如何适当地与被访人交谈等内容,使培训更有效。1977 年,坎内尔(Cannell)、奥克森伯格(Okesenberg)和康弗斯(Convers)进行了降低访谈员之间差异的尝试,他们把专门写好的自我介绍和一些鼓励语句交给访谈员,让他们在与被调查人交谈时使用,以减少调查员与被调查人员之间无结构的交流。如今,访谈员的培训和监督对保证资料质量的重要性已得到了充分的证明(Billet & Loosveldt,1988;Fowler & Mangione,1990)。

数据收集方式

直到 20 世纪 70 年代,大多数学术性的和政府的调查都是

采用面对面的访谈方式,并由普通的访谈员进行。当拥有电话的人在美国越来越普遍时,电话访谈便成为了数据收集的主要方式。最新的数据收集方式是利用互联网。目前由于很多人缺乏上网的条件,因此它的使用仍十分有限,但是,上网条件发展得十分迅速,毫无疑问,互联网很快就会成为主要的数据收集方式。

整体调查设计

在20世纪50年代,良好的研究实践原则从各方面都有了相当的发展。可以理解的是,随着科学技术的发展,研究的过程和手段也发生了变化。在有些情况下,对一个特定目的的研究而言,究竟什么样的数据收集方法才是最好的这一问题,我们还缺乏研究。即使在我们已经知道什么是最佳方法时,这些方法在实际使用时,其实施过程,在质量上也仍然会有这样或那样的差异。

导致调查质量不稳定的原因有很多。对某些调查而言,主要的问题是数字不精确。但在许多情况下,缺乏经费和合格的工作人员,以及缺乏方法论方面的知识导致了低质量的实践。缺乏严格的概率抽样以及问题的措辞不够标准化,可能也是导致调查质量不稳定的另外两个原因,但对这两个原因仍存在着一些争议(见 Converse,1987;Groves,1989;Suchman & Jordan,1990;Turner & Martin,1984)。但必须指出的是,调查研究设计中较为普遍存在顾此失彼的倾向,即对上述调查研究中三个基本方面的某些方面予以精心考虑的同时,却忽略了另外一些方面。本书的一个关键的取向是,把三个基本方面有机地联系在一起,形成高质量的研究设计,我们称之为整体调查设计的视角。

调查都要涉及一系列决策,而这一系列决策都会对调查统计值的准确性或精确性产生无形的影响——不是提高,便是降低。一般讲,那些导致"提高"的决策,都需要较多的经费、时间或其他资源。因此,研究设计无疑将涉及一组决策,以优化资

源的使用。优化设计与调查的三个基本方面都是密不可分的。

就抽样而言，关键性的问题有以下几个：

- 是否选择使用概率样本(probability sample)。
- 抽样框(sample frame，那些实际上有机会被抽样的人)。
- 样本大小。
- 样本设计(用来抽取人群或家庭样本的具体方案)。
- 回收率(在抽取的人中实际收集到了数据的人数的百分比)。

就问题设计而言，研究人员必须首先确定，问题的信度和效度究竟可以在多大程度上借鉴以往的有关文献。其次，研究人员还必须确定如何使用和求教于问题设计方面的专家。此外，他们还必须确定他们准备在试调查和问题评估方面投入多少资源。在调查员方面，研究人员可就有关调查员培训的方法、时间和内容，以及访谈质量的监控手段等有所选择。在设计决策中，一个将贯穿所有方面的决策是有关数据收集方式的决策。研究人员究竟采用邮寄、个别访谈还是其他方法收集资料的决定，不仅关系调查的费用，而且也会对调查质量产生影响。

将上述所有方面联系在一起，即我们所谓的总体调查设计(total survey design)，它从两个方面避免了以往各种研究设计的种种弱点，从而最大限度地降低了数据的误差。首先，在以往的某些研究中，我们发现，研究人员往往只注意调查中的某一两个方面，如样本容量、成功率等，并以此来评估数据的质量。而总体调查设计则要求研究人员注意上述与调查有关的所有方面。如果在调查设计的任何一方面中有了重大的疏漏或缺陷，调查工作的其他方面的大量投入就会失去效力。例如，如果提问一些被调查人不能给予准确答案的问题，一个以最大限度减少抽样误差为目的的规模很大的样本就会失去意义。类似的甚至更常见的情况还有：如果样本设计不严密，或者回答率小到使样本失去了代表性的程序，或者调查员缺乏训

练和监督,那么即使调查获得的回答者的数量再大,也不能增加调查的信度。

对调查研究的设计者和使用者而言,总体调查设计法意味着,在试图对调查的质量和某一特定数据集作评估时,必须考察与调查有关的所有方面和每一个环节,不要只注意其中的某些方面和某些环节。

本书的目的和目标

本书将对任何调查项目都不可或缺的重大决策问题展开讨论。任何决策都不是唯一的。根据不同的决策所作调查,其调查估计值的误差量和可信度也不尽相同。适当的研究设计给我们提供的一组程序,将使我们得以实施一个高质量的调查。我们将尽可能结合那些研究人员一定会遇到的实际问题,以及最为紧要的理论方法论上的问题来展开我们的讨论。许多实际数据收集工作中出现的问题,究其原因,往往是由于在实际调查中没有严格执行研究设计规定的程序,而非缺乏对研究设计的一般理解。换言之,问题在于"行",而非"知"。

限于篇幅,本书在讨论调查问题时必须有所选择。因此本书将要讨论和已经讨论的主要题目是:抽样、问题设计和有关调查员问题的一些研究等。那些要承担实际调查研究项目的读者则应阅读更多的有关著作。尽管必须指出的是,阅读本书及其他有关著作,都替代不了向那些在调查设计和实际操作方面既有深厚的方法论功底,又有丰富的实际经验的专家学习和请教,但是就本书本身而言,它也有着不可替代的重要作用,即它将在误差产生原因和调查数据的收集在方法论上所涉及的领域,给读者一个系统而全面的认识。

这些知识对许多人来讲已经足够,且弥足珍贵。当然对于那些使用由别人收集的资料的社会科学家来讲,则应对误差产生的原因有更进一步的了解。依此类推,那些使用以调查为基

础统计数据的读者,则应对数据收集的过程有更进一步的了解。本书还回答了那些使用数据的人必须提出且应得到回答的问题。此外,本书还为那些将进行招标或委托他人进行调查的人,就如何对调查的质量进行评估提供了一个简要而全面的方法。总之,本书旨在为那些可能成为调查研究的设计者或使用者的读者提供一些基本知识和方法,同时也为那些即将进行资料收集的人提供一个好的开端。

第2章 抽 样

样本能在多大程度上代表总体是由抽样框、样本大小以及抽样程序的特定设计决定的。如果使用的是概率抽样程序,样本估计值的精确度是可以计算出来的。本章对各种抽样程序以及它们对样本代表性和样本估计值的精确度的影响做了介绍,而对其中两种使用最为普遍的从总体中抽样的方法,即区域概率和随机数码拨号抽样做了比较详细的介绍。

有时,资料收集的目的不是为了获得有关总体的统计数据,而是要用更概括的方式描述一群人的特点。有时,新闻工作者、产品开发者、政治家等只需要大略了解人们的感情态度,而不太注意数字描述的准确性。如果研究者要进行小规模试验性研究以测量人们具有某种思想观点的广泛程度或者貌似一致而实则不同的程度,那么那些便于找到的人(如朋友、同事)或自愿参与者(如杂志调查的应答者、电话公开的人们)就会显得很适用。并非所有的资料收集都需要严格的概率抽样调查,然而在多数情况下,进行调查的目的是为了得到有关总体的统计数据。本章关注的就是以产生统计数据为目标的抽样,本章介绍的抽样方法,用于在需要产生适用社会科学的各种统计方法所需要的数字时。虽然本章介绍的很多一般原理同样适用任何抽样问题,但是着重讨论的是人的抽样问题。

评估样本的方法不是通过结果和样本的特点,而是通过考查选择样本的过程进行的。以下是样本选取的三个关键方面:

- 抽样框是在抽样方法选定之后,有着相同的被选择机会的人的集合。从统计学上讲,样本只能代表包含在抽样框中的总体。设计面临的问题之一,就是如何更好地使抽样框与所要描述的总体保持一致。
- 必须用概率抽样方法来确定包含在样本中的单独的个体单位。每个人必须都有一个由抽样方法确定的,已知的选择的机会。如果研究者的主观意愿性,或回答人个人的特性,如比较便于联系或有较高的参与调查的愿望等影响了选择的机会,那么我们便失去了评估样本对总体的代表性好坏的依据,因而使那些通常用来计算样本估计值的置信区间的方法也无法使用。
- 样本设计的细节,诸如样本的大小以及选择单元的特定步骤等,都会影响到样本估计值的精确性,即影响到样本能在多大程度上近似地反映总体的特点。

这些抽样操作过程的细节,以及能从被抽取的样本上获得

资料的比率,构成了评估一个调查样本质量所需的依据。本书第 3 章中将对回答率问题进行讨论,此外,还将简要讨论定额抽样(quota sampling),它是一种常用的,对概率抽样进行修正的抽样方法,它产生的是非概率(nonprobability)样本。本章将讨论抽样框、概率抽样方法,以及几种最为常用的抽取人的策略。对此有兴趣的读者可以在基什(Kish,1965)、萨德曼(Sudman,1976)、卡滕(Kalton,1983)、格罗夫斯(Groves,1989)、亨利(Henry,1990)和洛尔(Lohr,1998)的著作中获得更多的相关知识。尽管准备进行调查的研究者经常会被建议去咨询抽样统计学家,但是,本章还是努力让读者熟悉一些在评估一个调查的抽样时他们应当予以关注的,以及可能会遇到的问题。

抽样框

任何样本选择程序都会给某些个体一个入选样本的机会,而将其他个体排除在外。在选出的人中,那些有入选样本机会的人构成了抽样框。评估样本质量的第一步就是明确抽样框。一般讲,大多数抽样方案无非是以下三个类型中的一种:

- 抽样是在一张比较完整的研究总体的个体清单上进行的。
- 抽样是在一群比较固定的到某处或做某事的人(例如:接受医生治疗的人或是参加某个会议的人)中进行的。在这种情况下,抽样进行之前是没有清单的,清单的产生和抽样过程是同步的。
- 抽样是分成两个或多个阶段进行的,其第一阶段是对某些有关单位而非最终要抽取的个体进行选择。通过一步或几步,实现初级单位的抽样,并逐渐形成一份个体的(或其他抽样单位的)清单。最后,再以这些清单为基础,抽取最终的样本。诸如这样的抽样方法中,最常见的一种是在预先不知道户内居住者的情况下,先选择住户单位(HUs),

以此作为选取居住在这些户内的人的样本的第一阶段。这些多阶抽样方法,我们将在本章的后面部分进行详细的介绍。

研究者应该评估抽样框的以下三个特点:

1. **完整性**:一个样本所能代表的,只是它的抽样框而已,而所谓抽样框就是那个有被实际选取机会的人的总体。大多数抽样方法总会将一些人排除在研究的总体之外,例如,以住户为基础的样本就排除了那些住在诸如集体宿舍、监狱、疗养院等集体宿舍的人,也排除了那些无家可归的人。常见的可以获得的人员名单,如有驾照的司机、登记的选民、房屋所有者等有比较严重的排他性。虽然它们已经包括了某些总体中的大部分,但也排除了总体中的某些重要的有独特特征的部分。一个典型的例子是:公开出版的电话号码簿就排除了那些没有电话的人、要求不公开其电话号码的人以及在最新电话簿出版后才安装电话的人。在一些中心城市,几乎有50%的住户会因使用电话簿作为抽样框而被排除在研究总体之外。在这些城市里,用电话簿抽取样本,只对大约一半的总体有代表性,并且我们可以断定,这被排除在外的一半会在很多方面不同于另一半。

电子邮件地址抽样提供了另一个典型的例子。对于有的总体,如在公司或学校的人,使用电子邮件就很方便。而对一般住户总体使用电子邮件地址的抽样办法,就会排除很多人,产生的样本就会在许多重要方面与总体有很大的不同。

评估任何抽样设计的关键,是确定有被抽取机会的个体占总体的百分比,及其与没有被抽取机会的个体的差别程度。研究者会经常面临两难的选择:一个是简单便宜但会排除了一些人的方法;另一个是花费多但抽取的样本能更全面代表总体的方法。如果研究者想从已有的名单里抽样,特别重要的是要仔细评估这份名单以确定它是如何编制的、它的增删是何时及如

何进行的，并要确定可能被排除在名单外的人员的数量和特点。

2. **选择的概率**：有可能计算出每个被抽取的人的选择概率吗？从一个一年的就医记录中抽取个体的方法，将会使那些一年内就医数次的个体，比一年内只就医一次的个体有更高的选择机会。抽样设计没有必要给予抽样框中的每个成员相同的被抽取的概率。假如每个个体都在清单中出现一次，且只出现一次时，情况便是如此。然而，重要的问题在于，研究者应当能够算出每个被抽取的个体的选择的概率，这一点可以在选取样本时，通过对清单的考察来实现。有时，选择概率的确定也可以在数据收集时进行。

在上面提到的通过就医次数来抽取病人样本的例子中，如果研究者询问被抽取的病人一年就医的次数，或者研究者能获得被抽取的病人的就医记录，便有可能在分析的时候，考虑到选择机会上存在的差异，对数据进行修正。但是，如果研究者无法知道每个被抽样的人的选择概率，也就不可能准确估计样本统计值与其从中被抽取的总体之间的关系。

3. **有效性**：在有些情况下，抽样框包括了一些研究者不想抽取的单位，如果在数据收集时我们便能从中识别那些合格的个体，那么抽样框便不会过于庞杂。因此，抽取住户中老年人样本的最好的方法是先抽取一个各种类型的住户的样本，再确定选出的住户中是否有老人；并将那些没有老人的住户排除出去。随机拨号抽样是以抽取电话号码（有些并未使用）作为抽取拥有电话的住户的样本的方法。对这些设计而言，唯一存在的问题是它们是否合算。

因为从样本中进行概括的能力是由抽样框所决定的，所以在报告调查结果时，研究者必须告诉读者哪些人有被选择，或

没有被选择的机会，以及我们对选择机会的把握程度。此外，研究者还必须告诉读者，那些没有选择机会的人究竟有什么独特之处。

选取一阶样本(one-stage sample)

在抽样框或获得样本的途径确定之后，研究者接着要面对的问题是，如何选择那些将要被包含在样本中的个体。下面几节内容将讨论抽样者抽取样本的各种典型方法。

简单随机抽样(simple random sampling)

从某种意义上来说，简单随机抽样是总体抽样的基本类型。最基本的计算样本统计值的方法都假定抽取的是一个简单随机样本。简单随机抽样类似于从一顶帽子中抽取一个样本：每次有一个总体成员被选出，一个单位一旦被选出，便没有机会被再一次选到。

从操作上来讲，简单随机抽样需要一份有编号的总体的清单。为了简便起见，我们假定总体中的每个人都会在清单中出现一次，且只出现一次。假如一份清单上共有 8 500 人，而我们要选取一个 100 人的简单随机样本，那么选取的过程可能就很简单了。对清单上的人从 1 到 8 500 进行编号，然后使用电脑、一张随机数表或者其他产生随机数字的工具，就可以在同样的范围内产生 100 个不同的数字，与数目相对应的 100 人就组成了人数为 8 500 的总体的一个简单随机样本。如果清单是一个计算机数据文件，我们就可以先把清单上的编号进行随机排列，然后选择重新排序后的清单上的前 100 个人，这样我们将会得到一个等价的结果。

系统抽样(systematic samples)

如果清单很长，所有的抽样单位都未先编好号，或者不能

方便地用电脑操作进行编号,那么上述简单随机抽样的方法用起来就会十分困难。在这种情况下,可以使用另一种抽样方法,即系统抽样。系统抽样与简单随机抽样具有相同的精确性,并且操作方便。另外,系统抽样便于使分层的优越性体现出来(有关这一问题,我们将在下面的部分进行讨论)。

如果要从一个名单中选出系统样本,研究者首先要决定名单上条目的数量,以及将从清单中选取的基本单位的数量,然后用后者除以前者,便可得到一个分数。假如名单上有 8 500 人,样本只需 100 人,名单人数的 100/8 500 会被包括进样本(即从每 85 个人中选出 1 个)。为了选取一个系统样本,我们先在样本间距里选择一个随机数作为起点,在本例中就是从 1 到 85 这些数字中的任意一个数字。随机起点保证了选择过程的随机性。从那个位于随机选出的位置的人开始,研究人员在名单上每隔 85 个人选出 1 个,如此这般连续进行下去,直至选够 100 个人为止。

大多数统计学著作都认为,如果名单是按某些特征排序,或者存在某种循环的模式时,最好不要使用系统样本。因为,这时依据随机起点的不同,样本的排序特征或循环模式,将会对样本产生不同的影响。有一个极端的例子,在一个夫妇俱乐部的成员名单里,如果丈夫的名字总是排在妻子前面,那么只要是任何偶数的间距,总会产生一个只有一种性别的系统样本。在审查一个可能的抽样框时,我们必须注意这样一个问题:由某一个随机起点产生的样本,其影响样本结果的方式,是否会系统地不同于其他随机起点产生的样本。在实践中,大多数名单或抽样框不会产生不适合于系统抽样的问题。如果这样的问题果真出现了,只要调整名单的排序或抽样间距便可使问题迎刃而解。用这样一种策略得到的系统样本,至少可与简单随机样本相媲美。

分层抽样(stratified samples)

在抽取简单随机样本时,每次抽取都是独立的,不受先前

已经发生的任何抽取的影响。这样的过程导致的结果是,样本的任何一个特征,与它从中抽取的总体偶尔有所不同。一般说来,在数据收集之前,我们对单个总体成员的特点知之甚少。不过,总是会有一些总体的特点可在抽样时就能得到确认。在这种情况下,有可能重新安排抽样程序以减少常规的抽样变异,以便形成一个比简单随机样本更接近整个总体特征的样本。达到这个目的的过程被称作分层。

例如,某研究者有一个大学生名单,其中姓名是按字母排序的,名单里不同班级的学生名字混杂在一起。如果名单标明了每个学生所属的班级,就有可能将名单排序:把大一学生排在最前面,接着是大二、大三,最后是大四的学生,并且所有的班级都组合在一起。如果抽样设计要求抽取一个包括名单中的 1/10 学生的样本,那么重新排序后的名单就能保证,样本中有 1/10 的大一学生和 1/10 的大二学生,以此类推,其余年级的学生也为 1/10。如若不然,无论用简单随机抽样还是系统抽样,从按原来的字母排序的名单抽取的样本,其中大一学生的比例将取决于常规的变异性,可能略高于或略低于他们在总体中的比例。而提前分层则可确保样本中每个年级的比例刚好与它们在整个总体中的比例相同。

现在来考虑一下评估学生平均年龄的问题。由学生组成的班级几乎可以肯定是与年龄有一定关系的。虽然在抽样框中注意班级的代表性,可以在一定程度上抑制样本平均年龄偶尔出现的与总体差异,但是由于抽样过程的原因,在样本估计值中仍将存在一定变异。几乎所有地理区域总体的样本都可以依据某些区域变量进行分层,这样做可以使整个样本的分布方式与总体一致。国家的样本一般都以国家的不同地区,或城市、市郊和乡村进行分层。分层只会增加与分层变量有关的变量估计值的精确度。然而,因为一定程度的分层相对来说是比较简单易行的,加之它不会降低样本估计值的精确性(只要所有层次中的选择概率是相等的),因此,在通常情况下,我们总是希望样本设计能具有这样一个特点。

不同的选择概率(different probabilities of selection)

有时分层被用作改变总体中各种子总体选择率的第一步。在各个层次的选择概率一致时,一个占总体的10%的群体,也会构成样本的10%。如果研究者想从占总体10%的子总体中选出至少100人的样本,那么简单随机抽样法需要一个总数为1 000人的样本;如果研究者打算将子总体的样本容量扩大到150人,那么就需要追加500人到总的样本中,使样本的总容量达到1 500人,这样才能使总样本的10%等于150。

显然,有时以这种方式增加样本不仅费时费力,而且也不怎么有效。在后一个例子中,即使研究者对其他子总体的样本量是满意的,但为了增加50个我们希望增加的调查对象,在抽样设计中却不得不增加450个我们不希望增加的调查对象。因此,在有些情况下,合理的设计方法是让某些子总体比其他的子总体拥有更高的选择率。

例如,假设研究者想在一所男生仅占20%的大学里对男女学生进行比较,并希望男学生在样本中的人数不少于200个。这样,500人的样本中应该有100个男生。但是如果能事前区分出男生,就可以将男生的选择率提高为女生的2倍。以这样的方式抽样,我们不用为了增加100个男生,再追加500个被调查人。另加的100个男学生仍基于原来的500个人的样本,只不过样本中男生的总数变成了200。这样在进行男女生比较时,我们将会有由200个男性被调查人和300个女性被调查人的样本提供的精度。为了将这些样本合并在一起,研究者应给予男性被调查人的权值为女生的一半,因为男生的选择概率是女生的2倍。

在表2-1里,可以看到被调查者中有33.3%的是男生(600人中有200个男生)。但是采取了加权的办法后男生组成了加权后样本的20%,这就与他们在总体中的所占比例相等。

即使对所关注的子总体中的个体不能在抽样前了解清楚,有时也能应用这里提到的基本方法。例如,一般不可能在接触

前就获得确认了居住者种族的住户单位的清单。然而,在某些邻里地区,亚裔家庭比其他族裔的家庭居住更为集中,这样的情况并不少见。这种情况下,研究者就可以在亚裔占优势地区,用高于平均数的比率抽取亚裔家庭,以增加亚裔调查对象的数量。再次强调,在任何子总体与总体中的其他子总体的选择率不同时,为了使组合的或总的样本能产生关于总体的准确的统计值,就需要进行适当的加权处理。

表 2-1 不等概抽样示例

	女生	男生
总体的人数	4 000	1 000
在总体中所占百分比	80	20
抽样分数	1/10	1/5
样本中的人数	400	200
未加权的样本百分比	66.7	33.3
权(以调整选择概率)	1	1/2
加权样本人数	400	100
加权样本百分比	80	20

第三种方法是在与有可能成为调查对象的人接触后,依据从他们那里收集到的资料,对选择率进行调整。仍以上述大学生调查为例,假设事先不能查明学生的性别,研究者可以选择一个有 1 000 名学生的初始样本,让访谈者把每个学生的性别查明,然后对所有被选的男生(200)和一半的被选女生(400)进行访谈。其结果也会与刚才讨论过的表格一样。

最后,我们认为有必要谈一下,在不同的层使用不同的选择率的另一个技术上的原因。如果就测量的变量的变异程度而言,某一个群体与其他群体存在很大不同,那么对那个具有较高变异度的群体进行超比例的抽样,可能会有助于总估计值精度的提高。格罗夫斯(Groves,1989)很好地阐述这一基本原理,并提供了评价这些设计的效度的方法。

多阶抽样(multistage sampling)

当没有合适的总体成员名单,并且没有办法直接把握总体时,多阶抽样就成为一种很有用的抽样方法。

在缺乏直接的抽样的资料时,就需要采用一种将总体成员与某种能够进行抽样的分组联系起来的策略。第一阶段先对这些分组进行抽样,然后再设法编制选定的这些组的成员名单,以便在第二阶段(或稍后的)能从这些创建的名单中做进一步的抽样。在抽样术语中,抽样设计最后一阶段的分组通常被称为"群(clusters)"。在下面几节,我们将通过多阶抽样在三种最为常见的无法得到总体成员清单的场合的使用,来阐述多阶抽样的一般策略。

从学校里抽取学生样本

在从某城市的公立学校中所有注册在读的学生里抽出一个样本时,找不到诸如这样的个体的单独一个完整名单是常有的事情。然而有一种抽样框,却既是研究者可以得到的,同时也包含了期望的总体内的所有学生。这个抽样框就是该市所有公立学校的名单。因为研究总体中的每个学生都属于一个学校,且只属于这些学校中的一所学校,因此我们可以采用二阶抽样策略,选取一个完全可以接受的学生样本:首先选择学校(即群),然后从那些学校里选择学生。

假设有以下数据:

城市里有 40 个学校共有 20 000 名学生
希望的样本 = 2 000 = 1/10 的学生

那么将有如表 2-2 所示的四种抽样设计或方法,其中每一种都会产生 2 000 人的概率样本。

表 2-2 多阶抽样示例

	第一阶段的选择概率(学校)	×	第二阶段的选择概率(在所选学校里的学生)	=	总的选择概率
1) 选择所有学校列出所有学生,然后在每个学校选择 1/10 的学生	1/1	×	1/10	=	1/10
2) 选择 1/2 的学校,然后在其中选择 1/5 的学生	1/2	×	1/5	=	1/10
3) 选择 1/5 的学校,然后在其中选择 1/2 的学生	1/5	×	1/2	=	1/10
4) 选择 1/10 的学校,然后收集其中所有学生的资料	1/10	×	1/1	=	1/10

以上四种方法中的每一个都将产生一个 2 000 人的样本,且都给予该市的每一个学生相等的(1/10)选择几率。四种方法的不同在于:从上到下四种方法,所需经费越来越少,名单涉及的学校越来越少,需要走访的学校越来越少。与此同时,样本的精确度会因抽取了更少的学校和每所学校抽了更多的学生而有所降低。有关这样和那样的多阶抽样设计对样本估计值精确度的影响,将在本章后面几节进行更详细的讨论。

区域概率抽样(area probability sampling)

区域概率抽样因其广泛的适用性而成为多阶抽样策略中最常使用的方法之一。它可以用来对任何可以在地理上定义的总体进行抽样,例如,居住在某个邻里地区、城市、州或国家的人们。基本方法就是将所有总地区划分为穷尽而互斥的,有明确的边界的子区域。而这些子区域都是群,先抽取一个子区

域的样本,然后再设法编制选出的子区域内的住户单位的清单,并抽取清单上列出的单位的样本。最后一步,或将选出的住户单位内的所有人都包含在样本内,或设法编制他们的名单,也对他们进行抽样。

这种方法适用于丛林、沙漠、人烟稀少的乡下或大城市的中心区。抽取这种样本的具体步骤可能十分复杂,但是,通过以城市街区为初级子区域单位抽取城市总体样本这一例子,我们可以比较清楚地阐明它的基本原理。假设我们有以下数据:

一个城市有400个街区
共有20 000个住户单位住在这些街区里
希望的样本=2 000个住户单元=1/10的所有住户单位

如果我们已经掌握了这些信息,便可用与从学校抽取学生样本相似的方法去抽取住户样本。抽样的第一个阶段,选定街区(即群),如表2-3所示。在第二个阶段,列出被选定街区里的住户单位名单,从名单里抽取样本。接下来介绍两种抽取住户单位的方法。

表2-3 区域概率抽样示例(一)

	第一阶段的抽样概率(街区)	×	第二阶段的抽样概率(选取街区的住户单元)	=	总的抽样概率
1)先选取80个街区(1/5),然后再选取这些街区里1/2的单元	1/5	×	1/2	=	1/10
2)先选取40个街区(1/10),然后再选取这些街区里所有的单元	1/10	×	1/1	=	1/10

与学校抽样相类似，由于第一种方法比第二种所选定的街区更多，因此，成本更高，在给定的样本容量下，它产生的样本估计值精度更高。

这些抽样方案中没有一个考虑到了第一阶段的组群的大小(例如，街区或学校的大小)。大的学校和大的街区的选择率与小学校和小街区相同，如果，在最后一个阶段，每个被选出的组群都取一个固定的分数，那么与小的学校或街区相比，大的学校或街区的访谈的数量将更多。在第二阶段抽取的样本，其大小(群的大小)就会有很大的差别。

如果有关于第一阶段抽取的组群大小的资料，那么对这些资料予以运用往往是很有益的。如果在最后一步选择中，选取的单位数在所有的群中大致相等，那么这样的样本设计将产生更精确的估计值。这种设计方案的优点还在于能更容易计算出抽样误差，并且更容易预见样本的总容量。为了产生同等容量的群，在第一阶段抽样时，单元的选择应与它们的大小成比例。

接下来的例子展示了与街区规模成比例地对街区进行抽样的方法。它是以区域概率法作为对住户单位(公寓或个人住房)进行抽样的第一步。这种方法也适用于学校抽样，只要把学校当作街区就可以了。

1. 确定在抽样的最后阶段要抽取的住户单位的数——整群规模平均数，例如，我们取它为 10 个。
2. 估计第一阶段单位(街区)内的住户单位数。
3. 给街区排序，以便把地理位置邻近或相似的街区排在一起，这种高效的有层次抽样，将提高样本的质量。
4. 产生所有街区住户单位的估计累计数，最终将形成表 2-4。
5. 确定群的间距。如果要从 10 个住户单位中选出 1 个，并且在每个选定的街区中选择一个由 10 个住户单位组成的群，就需把群的间距定为 100 个住户单位。换种方

法，我们不从每10个住户里选出1个，而从每100个住户里选出10个，选择率是一样的，但是使用的模式却是“成群的”。

表2.4 区域概率抽样示例(二)

街区号	住户单元估计数	家庭单元的累计数	得分（随机起点＝70，间距＝100个家庭单元）
1	43	43	—
2	87	130	70
3	99	229	170
4	27	256	—
5	15	271	270

6. 在从1到100（本例的间距）取一个随机数作为起点，然后按照累计数系统地进行处理，将第一阶段选出的初级单元（街区）一一标明。在本例中，选中的随机起点(70)就跳过了第1街区（尽管在100次选择中，它有43次被选中的可能），第70个住户单位在第2街区，第170个住户单位在第3街区，第270个住户单位在第5街区。

然后我们要编制那些被选中的街区（第2，3，5街区）中住户单位的清单，通常都是派人到这些街区，实地走访进行编制。清单编好后，便要从这些清单中选出住户单位。如果我们能保证每个街区规模的估计值都是准确的，那么研究者只要用简单的随机抽样或系统抽样，在每个选定的街区选出10个住户单位就可以了。一般认为系统抽样是最好的，因为由此选定的单位会分布在整个街区。

一般讲，第一阶段抽样单位（例如街区）规模的估计值或多或少总是会有一些误差。我们可以通过计算从街区中选择住户单位的比率来修正这些误差，以街区2为例，其计算公

式是:

$$街区住户选择率 = 群规模的平均数/街区住户单位的估计数$$
$$= 10/87 = 1/8.7(在第2街区里)$$

本例中,在第 2 街区,我们可以在每 8.7 个住户单位里抽取 1 户;在第 3 街区,每 9.9 个户里取 1 户;在第 5 个街区,每 1.5 户取 1 户。如果某街区规模比预想的大(例如,进行了新的建设),抽取的户数就不止 10 户;如果它比预期的小(例如,进行了拆迁),抽取的户数就少于 10 户;如果与我们预想的一样(例如,第 2 街区中有 87 个户),我们就选 10 户(87 ÷ 8.7 = 10)。通过这种方式,就可以在抽样的过程中自动修正最初对街区规模评估的误差,同时使得所有街区中的住户单位都有同样的被选择的机会。不管是估计的还是实际的街区规模,任何住户单位被选择的概率都是 1/10。

区域概率抽样法可以用来对任何可以做地域划分的总体进行抽样。虽然区域越大,步骤就会越复杂,但是基本路数都是一样的。要记住以下的关键步骤:

- 所有区域都必须给予某种被选择的机会;将那些预计没有任何抽样单位的区域与其相邻的区域合并以确保选择的机会;注意区域内是否有新建筑或估计值是否可能有错误。
- 街区(或其他类型的地区)的选择概率与该街区内住户选择概率的乘积,在所有的街区都应该保持不变。

最后,即使是细心的实地制表人也会漏掉一些住户单位。因此,一种值得效法的做法是,在数据收集时,在清单上补上那些原来遗漏的单位,并做记号。

随机数码拨号

随机数码拨号(ramdom-digit dialing, RDD)为我们提供了

一种备择的，为了抽取样本户中的人而抽取一个住户的样本的方法。假设有 6 个电话局覆盖了 20 000 个住户单位。研究者可以抽取一个 10% 的有电话的住户的概率样本，方法如下：

- 这 6 个电话局共有 60 000 个电话号码（每个电话局 10 000 个）。从这些数中选出 6 000 个（即 10%），就是从每个电话局抽取 1 000 个随机产生的四位数。
- 拨打所有这 6 000 个号码。并非所有的号码都是住户电话号码，实际上，这些号码中还有许多是未曾投入使用的、接不通的、暂时未在使用的或是商户电话号码。因为可能运行于该地区的电话号码的 10% 已经被拨打过了，所以通过拨打这些样本电话号码，最多有可能联系上约 10% 的有电话的住户。

这就是最基本的随机数码拨号抽样法。此法最明显的不足就是有大量没有结果的拨叫。在全国范围内，居民住宅的号码不足 25%，城镇地区这一数字约为 30%，而在农村只有 10%。瓦克斯伯格（Waksberg，1978）发明了一种特别的随机数码拨号法，这种方法利用了电话号码的分配是成组的这一优点。每个含 100 个电话号码的组，都由 3 个地区号码，3 个电话局号码和 2 个附加码（区位号码-123-45 × ×）定义。首先，通过拨打组样本中的一个随机号码，对号码进行初步筛选，然后只拨打发现了住宅电话号码的那些组的附加随机码。采用这种方法，可使拨到住宅电话号码的比率上升到 50% 以上。在这个设计中，100 个电话号码构成的组就是“群”。

近些年来，大多数调查机构开始使用名单辅助的随机数码拨号法。随着电脑技术的发展，公司可以编辑电子电话簿，这些电子电话簿每 3 个月更新一次。一旦所有的电话簿都是以电子文件的方式存在，就可以搜寻到至少有一个公开（出现在电话簿上的）住宅电话号码的所有群（区域码-123-45 × ×）。这些公司就可把所有至少有一个公开的住宅电话号码的群的群内所有可能的电话号码构建一个抽样框，然后利用这个抽样

框进行抽样。这种方法有两个优点:第一是瓦克斯伯格方法里所要求的电话号码的初步的筛选就不再需要了,因为编好的抽样框已经完成了这项工作;第二是使用这个抽样框选好的样本不再需要分组了,通过使用包含了所有的住宅电话号码的群的抽样框,就可对电话号码进行简单随机抽样了。这种新方法比以前的方法更加经济有效。其局限在于,在群中,没有被列出的住宅号码就没有被抽中的机会。布里克、瓦克斯伯格、库尔普和斯塔尔(Brick,Wakesberg,Kulp,Starer,1995)估计在美国平均有4%的有电话的住户未刊登在电话簿上。列普克夫斯基(Lepkowski,1988)对为了抽取住户而实施的各种电话号码抽样法进行了详细的总结。

另外,在使用随机数码拨号抽样法时,还应该注意一些问题。首先,它的使用价值是建立在大多数住户有电话的事实基础之上的。在全国范围内,大概只有5%的住户没有电话,但是在一些地区,特别是中心城市或乡村地区,丢失率可能更高。桑贝里和马西(Thornberry & Massey,1988)对那些有和无电话的住户之间存在的差别做了进一步的分析。

随着手机使用率的上升,家庭座机使用率下降,RDD 抽样将面临一个问题:目前采用的 RDD 抽样法,只关注住户座机,而完全忽略了手机用户。当然,同时对这两种电话用户同时进行抽样是可能的,但是假如将手机号码也加入到抽样框中,那么抽样、资料收集和调查后的加权就会变得十分复杂。

对 RDD 抽样面临的另一个挑战是,在那些电话局号与地区边界不对应的面积较小的地区,找到一种能对那里的人进行抽样的方法。除非我们能确信电话局的局号码与研究区域相一致,否则访谈员就需要回答者告诉他们,自己居住的地方是否住在调查区内。在回答者居住在一个面积很小,定义模糊的地区,如一个邻里地区时,即使问了回答者人这样的问题,我们仍然还是难以确定他们居住的地方是否在调查区内。

像其他特定的抽样方法一样,对所有的调查而言,随机数码拨号法并不一定是一种最佳设计。对其赞成和反对的观点

将在第4章进行详细讨论。无论怎样,在过去的30年里,作为抽样方法之一的随机数码拨号法的引进,为调查研究水平的提高作出了重大贡献。

选择回答者

区域概率样本和随机数码拨号样本都是对住户单位样本,因而,接下来的问题是我们应该对户中的哪一个人进行调查。

最佳的决策取决于我们要收集的信息是什么样的。在某些研究中,要对住户及其所有成员的信息加以收集。如果要求回答的问题很简单,那么户内任何成年人都可以作为回答人;如果要回答的问题很特别,那么研究者可能就会要求户中的知情者来回答。例如,在国家健康调查里,那个"最了解家庭健康状况"的人将被作为那些涉及家庭所有成员的问题的回答者。

但是,有很多事情,很多人都只能说出关于他(她)自己的事来。几乎所有的研究人员都认为没有任何人可以说出他人的感受、观点和知识。还有一些通常只有自己才能准确说出的行为和经历(例如:吃喝的东西、购买的物品、看见或被告知的事物)。

当研究包括的变量只有通过自我讲述方式得到才比较合适时,抽样过程除了选定家庭之外,还必须从这些家庭中抽取特定的个人。一种方法就是与家庭中的每个符合条件的人进行访谈(此时,不进行任何抽样)。因为户内的同质性,及担心一个回答人可能会对后面的回答人有影响,所以最常用的方法是在每户只指定一个回答者。很显然,选择那些恰巧接电话或开门的人作为回答者是非概率的和一种有偏倚的选择个体的方式;访谈员的和回答人的主观随意性,以及个体的可得性(它与工作状况、生活方式和年龄相关)等,都会对什么人将会成为回答人有影响。概率抽样的重要原则是选择是通过某种机遇,或用以指定特定的人的随机程序进行的。在户内产生选择回答人的概率程序涉及以下三个步骤:

- 确定户内有多少合格的回答人(例如,有几个 18 岁或 18 岁以上的人)。
- 采用一种对所有的户都相同的方式,对这些合格者进行编号(例如,按降序排列年龄)。
- 有一个客观的选择回答人的程序。

基什(Kish,1949)编制了用一套沿用至今的随机化表格来指定回答人的详细步骤。在访谈有电脑辅助时,用电脑在合格的户内成员中选出一位回答人是一件很容易的事。这个程序的关键所在是不受主观随意性的干扰,和所有选出的户内的合格成员都有一个已知的(非零的)选择概率。格罗夫斯和利伯格(Groves & Lyberg,1988)在他们的著作中,对几种简化回答人选择程序的策略做了评论。

在只需对户内的一个人进行调查时,就会引起选择率上的差异。如果一个成年人居住在一个单成人户中,显然,在这一户被选中时,这个成年人便成为了回答人。相比之下,一个居住在有 3 个成人户中的成年人被选做回答人的机会只有一人户的 1/3。每当一组可以确定的个体与以其他组的个体以不同的选择率被选取时,我们必须采用加权对之进行修正,以使那些过度抽取的人在样本统计值中不会有过度的代表。在本章的前面的例子中,当男生的选择率是女生的两倍时,他们的答案就得乘以权值 0.5,这样他们的样本加权比例与总体中的比例相同。在从合格人数的不同的户中选择一个回答人时我们也可以使用同样的方法。

对每个家庭只选择一个回答人的影响进行修正的最简单方法是,用户中符合条件的人数给每个答案加权。因此,如果有 3 个符合条件的成年人,那么权值为 3;如果有 2 个成年人,权值就为 2;如果只有 1 个符合条件的成年人,权值就为 1。如果加权方案是合适的,那么对于所有的回答者来说,抽样概率乘以权值的积都是一样的。

还有一些变量与家庭中成年人的数量紧密联系。例如,结

婚的人比单身的人更有可能会住在至少有两个成年人的户里。已婚者与单身者在很多方面都有所不同,因此,假如数据未进行加权处理,那么任何与婚姻状况相关的估计值都将是不准确的。

也有一些变量与户内的成人数没有关系,在这种情况下,加权也不会对描述性的结果产生影响。当需要产生描述性统计数据时(例如均值或分布),一般通过加权来修正家庭和(或)成人不同的选择率。此外,为描述性目的而进行的加权对很多标准分析程序包来说是十分容易的。但是根据加权数据进行统计检验问题就会变得比较复杂。用于统计检验的估计值、回答关系到实际的答案数,及其他样本设计具有的一些特性,因此,假如使用的分析程序假设的观察数,多于或少于实际所有的观察数,加权就可能歪曲计算的结果。有时在一个调查设计很复杂时,想要准确地进行统计检验并不是一件容易的事。不仅如此,即使对某些相关分析而言,加权可能也未必都是必需的(Groves,1989)。这是对于恰当的设计和调查数据的分析之所以需要有抽样统计学家参与的又一例证。

样本估计值和抽样误差

我们之所以选择上面那些抽样策略给大家做介绍,不仅是因为它们的使用极其普遍,也因为它们阐明了可供我们选择的主要的抽样设计方案。一个概率抽样方案会在没有研究者或回答者的主观随意性的影响下,逐渐指定一组特定的住户或个人。研究者所能使用的基本工具就是简单随机抽样和系统抽样,它们都可以通过分层、不等的选择率和集群对其加以完善。抽样策略的选择在一定程度上取决于可行性和成本,当然,样本估计值的精确性也应在考虑之列。使用概率抽样方法的主要原因在于,它使我们能使用各种统计工具去评估样本估计值的精确性。在这一部分,我们将对这样一些样本估计值的计算

方法，以及它们是如何受样本设计特点影响问题进行探讨。

研究者常常对样本本身的特点不感兴趣，收集样本数据的目的是为了得出有关整个总体的结论。本章对于统计和设计问题的讨论，是在我们可在多大程度上确信样本特征精确地描述了整个总体这一背景下展开的。

许多常见的现象被用来解释概率理论。而其中最容易理解的可能就是掷一枚硬币。如果一个质量均匀的硬币被掷了 10 次，其结果不一定恰好是 5 个正面和 5 个反面，有时是 6 个正面，偶尔是 7 个正面，甚至还会出现 10 个正面。实际上，你可以把任何一个 10 次掷硬币的活动看作是许多可能样本中的一个。如果进行了一系列的 10 次掷硬币，在掷了 10 次后记下硬币正面出现的次数，再掷 10 次再记录下硬币正面出现的次数，一直这样做下去，这样就会产生一种分布。如果硬币的质量是均匀的，那么正面为 5 次的样本数将会多于任何其他正面次数不等于 5 的样本数。此时，也将会出现一个围绕 5 次朝上的分布，在这种分布中，处于分布两个极端的 10 次面朝上和 1 次朝上出现的频数最低。

虽然调查中有些误差来源于偏倚导致的数字的系统性失真，但是抽样误差是抽样随机的（因此不是系统性的偏倚）结果偏倚。在用概率抽样法选取样本时，我们可以计算出由于抽样而引起样本估计值偶然变化的幅度。

如果抽取的样本数为无限多个，那么样本的描述性统计量（例如，均值）的估计值就会形成一个围绕真的总体值（true population value）的正态分布。样本量越大、测量内容的方差越小，样本估计值就越紧密地成串集结在真总体值的周围，并且以样本为依据的估计值就会越精确。围绕真总体值的变差（源于样本和总体之间的偶然差异）被称为“抽样误差（sampling error）”。在给定的正常的样本变异性下，估计我们对样本估计值可具有的置信性的大小，是对来自调查的数字进行评估的一个重要组成部分。

样本设计的选择（尤其是选择涉及分层、集群或不等的选

择概率时)会影响一容量给定的样本的抽样误差的估计值。但是,通常描述抽样误差的方法,是先计算简单随机抽样中可能出现的抽样误差,然后再计算因背离简单随机抽样设计而产生的效应。因此,我们首先要介绍的是简单随机样本的抽样误差的计算。

简单随机样本中的抽样误差

本书不是关于抽样统计的教科书,然而,估计可以预期的来自某个样本设计的误差的量也是调查设计过程的基本组成部分。不仅如此,研究者照例也必须给读者提供了解应归结于抽样误差的指南。无论是出于求知目的读者,还是调查数据的用户,都应了解并理解这些指南。因此,对如何计算抽样误差这一问题的认识是了解整个调查过程的必要组成部分。

尽管同样的逻辑适用于根据一个样本计算的统计量,但是最为常见的样本调查估计值是均值或平均数。最为常用的描述抽样误差的统计量叫做标准误差(standard error)(或叫均值的标准误差),它是在样本量给定的情况下有限次地抽取一个又一个的样本时,形成的样本均值的抽样分布的标准差(standard deviation)。标准误差值一旦被估计出来,人们就可以说容量给定的和设计的样本均值的67%将落在真总体值±1个标准误差的范围内,诸如这样的样本的95%将会落在真总体值±2个标准误差的范围内。后面一个数字(±2标准误差)通常被称为样本估计值的"置信区间(confidence interval)"。

均值的标准误差的估计值的计算,是以方差和样本量为依据进行的:

$$\mathrm{SE} = \sqrt{\frac{\mathrm{Var}}{n}}$$

SE——均值的标准误差

Var——方差(样本离差的平方和除以 n)

n——样本大小

从样本调查中计算出来的一种最常见的均值可能是比值，即有某种确定特点或给出了某种回答的样本的百分数。它在表明为什么比例是双值分布的均值时是很有用的。

均值就是一种平均数，它的计算方法是将值和除以个案数：$\sum X/n$。现在假设只有 2 个值 0（否）和 1（是），某样本中有 50 个成员，当问到是否结婚时，有 20 个人说是，其他的说否。如果有 20 个“是”和 30 个“否”的回答，就可这样计算均值：

$$\sum X = (20 \times 1) + (30 \times 0) = 20; \frac{\sum X}{n} = \frac{20}{50} = 0.40$$

一个诸如有 40% 的回答人结婚这样的比例陈述，恰是一个关于 1/0 分布的均值的陈述，均值为 0.40。由比例的方差为 $p \times (1-p)$ 这一点，我们不难得知比例的标准误差的计算方法，即将 $p \times (1-p)$ 开平方即可。式中的 p 就是具有某种特征的比例（例如，在这一例子中有 40% 的人已婚），$(1-p)$ 是样本中没有某种特征的比例（即 60% 的人未婚）。

我们已经了解，均值的标准误差应该如下式所示：

$$\mathrm{SE} = \sqrt{\frac{\mathrm{Var}}{n}}$$

因为 $p(1-p)$ 是一个比例的方差，所以比例的标准误差计算公式变为：

$$\mathrm{SE} = \sqrt{\frac{p(1-p)}{n}}$$

在前面例子中，50 人中有 40% 的人已婚，其标准误差的估计值应该等于：

$$\mathrm{SE} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.40 \times 0.60}{50}} = \sqrt{\frac{0.24}{50}} = 0.07$$

因此，我们可以估计真总体数（整个总体中已婚比例）在

0.33 到 0.47(0.40 ±0.07)之间的概率是 0.67(即样本均值 ±1 个标准误差)。我们有 95% 的把握说总体的真值在样本均值的两个标准误差之间,即在 0.26 和 0.54(0.40 ±0.14)之间。

表 2.5 概括了不同容量和比例的样本的抽样误差,表中所提供的样本是通过简单随机抽样产生的,表里出现的所有数字代表了一个比例的两个标准误差。在给出特定的回答的样本比例已知(可以估计)的前提下,表格给出了不同样本容量的 95% 的置信区间。在本例中,50 个样本中若已婚比例的估计值为 40%,表格列出的置信区间在 0.14 左右,它与我们刚才计算得到的数字相同。如果 100 个人中已婚比例的估计值为 20%,表格显示,我们有 95% 的把握说真值在 20% ±8% 的区间内(即 12% 到 28%)。

表 2.5 不同抽样变异的置信范围*

	有某些特点的样本的比例				
样本规模	5/95	10/90	20/80	30/70	50/50
35	7	10	14	15	17
50	6	8	11	13	14
75	5	7	9	11	12
100	4	6	8	9	10
200	4	6	8	9	10
300	3	4	6	6	7
500	2	3	4	4	4
1 000	1	2	3	3	3
1 500	1	2	2	2	2

注:在样本报告的有某特征的百分比和该百分比依据的样本数均为给定的情况下,真总体数位于表格指出的 ± 数范围内的机会是 95%。

* 本表描述了由抽样引起的变异性。源于无回答或报告错误的误差没有反映在本表中。此外,本表假定使用的是简单随机样本。由于样本设计或其他方面的影响,估计值的变异性或许比本表列出的更大。

有关这一表格,我们认为读者应该注意几点:首先,我们不难看出增加样本量可以降低抽样误差。其次,我们也不难看到,在将一定数目的个案加到一个样本中可以降低抽样误差。

在样本量比较小的时候，较之样本量相对较大的时候，误差的降幅要大。例如，将 50 个成员加入容量为 50 的样本中，就会使误差明显降低。但是将 50 个成员加入到容量为 500 的样本中，它对增加整个样本估计值的精确度所起的作用几乎看不出来。

第三，我们还可以看到，抽样误差的绝对值在比例为 0.5 左右时最大，当具有某种特点的样本比例达到 0 或 100% 时，抽样误差的绝对值会下降。可以看到标准误差与方差有直接联系。当比例离 0.5 越远时，方差 $p(1-p)$ 就会变得越来越小。

第四，表 2.5 和它所基于产生的公式适用于简单随机抽样所产生的样本。一般总体的大多数样本不是简单随机样本，样本设计对计算抽样误差的影响程度，会因同一调查里的不同设计和不同变量而有所不同。在更多的情况下，表 2.5 会低估一般总体样本的样本误差。

最后，要强调的是，反映在表 2.5 中的变异性，所描述的是源于抽样而非源于总体中每个个体信息的收集的潜在误差。这个计算没有包括来源于调查过程其他方面的误差估计值。

其他样本设计特点的效应

前面讨论了简单随机抽样误差的计算方法。抽样误差的估计值会受不同抽样方法的影响。如果没有分层，系统抽样产生的误差会与简单随机抽样相同。如果每一层的选择率是一样的，那么对于同样容量的样本来说，分层样本的抽样误差比简单随机样本的要低，因为变量是随层而异的（平均而言）。

不等选择率（以不同的比率从总体中抽取子群）的设计，是为了增加超比例抽样的子群的估计值的精确度。

总体讲，不等选择率产生的整个样本的误差，高于等容量的简单随机样本，因为变量是随层而异的。

例外的情况发生在对含某些方差大于平均数的变量的层做超比例抽样时，这时，这些变量的总样本误差将低于等容量的简单随机样本。

集群产生的抽样误差将比等容量的简单随机样本高，因为，总体讲，变量在群中的同质性比总体的高。此外，最后一阶的群的规模越大，在通常情况下，它对抽样误差的影响也就越大。

一般情况下，设计特点对估计值精确度的影响是很难预料的。设计影响会由于研究的不同和同一研究中变量的不同而有所不同。为了阐明这一点，我们假设，我们选出的各个不同的街区只考虑每所房子的建筑类型必须相同，而不论它们是否由业主本人占用。这样，即使在一个街区的一个回答人报告说自己是住宅的主人，诸如这样的在那个社区进行的额外访谈，绝对不会给整个总体的房屋所有权比率问题增加任何新的信息。正因为如此，不管研究者对每个街区只调查一家还是调查20家，该估计值的信度其实都是一样的，它基本上都与进行访谈的街区的数成比例。另一个极端情况是，某街区里成年人的身高差异与整个城市成年人的身高差异相同。当某街区的回答人的同质性与总体相同时，整群抽样不会降低一个给定的身高样本估计值的精确度。因此，为了估计集群对抽样误差的可能影响，研究者应该注意群或分层所具有的特点以及要得到的样本估计值的种类。

样本设计对抽样误差的影响经常不受重视，经常会看到在整群样本设计时，报告的置信区间其假设却是简单随机抽样。想要提前知道设计效应的大小是十分困难的。正如所提到的那样，样本设计对抽样误差的影响就每个变量来说都是不同的。当样本设计不是使用简单随机抽样时，如使用整群和分层，这些计算尤其复杂。因为计算抽样误差的能力是运用调查方法的最重要组成部分之一，因此，让统计学家参与具有复杂样本设计的调查，以确保能计算并恰如其分地告知样本误差，是十分重要的。在评估样本误差时，对设计特征效应进行适当考虑的问题已经大大地简单化了，因为我们已经开发出来了做这样的调整的计算机分析程序包。

最后要指出的是，把样本设计的特点放在整个调查目标中

进行评估的做法是值得效仿的。整群抽样可以节约抽样(编制清单)和数据收集的成本。不仅如此,我们还发现,对许多变量来讲,整群样本一般不会过多地加大它们的抽样误差。一个或多个组群的超比例抽样经常是一个很划算的设计。像在本书提到的很多问题一样,对一个研究者来说,重要的是要意识到所做选择的潜在的成本和效益,并把它放到所有可供选择的设计中,结合调查的主要目的加以衡量。

样本的规模应该多大

对一个调查方法学家而言,在样本设计涉及的诸多问题中,最普遍的问题之一是样本的容量究竟应该多大。在提供答案之前,也许最好先谈谈三种普遍但不是很恰当的确定样本容量的做法。

一种普遍的误解是,一个样本的精确性在很大程度上依赖于包含在那个样本中的总体的分数——就是说诸如总体的1%、5%这样的百分数,或总体的其他百分数将使样本变得可信。迄今为止已经进行过的关于抽样误差的估计值的讨论,并没有考虑过包含在样本中的总体分数问题。前面的公式和表中得到的抽样误差估计值,可以通过乘以值$(1-f)$使其降低,其中:f=样本中所包含的总体的分数。

当研究者抽取的样本为总体的10%或更大时,这个调整对抽样误差的估计值有明显影响。然而,大多数样本即使再大也只包括了总体中很小的一部分,在这种情况下,样本中包含的总体的分数即使增加一点,也不会影响研究者从样本推断总体的能力。

我们也可以对这一原理做反向的思考。样本赖以抽取的总体的规模的大小并不会对样本描述总体的好坏程度产生实质性的影响。假设样本设计和抽样过程的所有方面都是一样的,一个由150人组成的样本对一个由15 000或1 500万人组

成的总体进行描述,可以得出同样有效的精确结论。与总的样本规模和其他设计特点相比(例如整群抽样),抽取的总体的分数对抽样误差的影响微不足道,在决定样本规模的时候,对其进行着重考虑是不必要的。

第二种不恰当的确定样本量的做法比较容易理解一些。有些人受到了所谓的“标准调查研究”的影响,他们据此引申出了典型的或恰当的样本量。于是,有的人会说全国性调查样本一般来说是1 500 人,或者说一个好的社区调查样本是500人。当然,参考有实力的调查者所认为的对特定总体来说最适宜的样本规模不失是一种聪明的做法。但是,像大多数其他的设计一样,样本量的确定必须是建立在具体问题具体对待的基础上,要以研究者所要达到的各种研究目标以及对研究设计的许多其他方面的考虑为转移。

第三种,也是特别需要提到的一种确定样本规模的错误方法,因为它可以在很多统计学的教科书上找到。方法是这样的:研究者先确定能允许的误差的容限或所要求的估计值的精确度,一旦明确了所要求的精确度,就可以简单地使用诸如表2.5 介绍的各方法,或对它们作了适当变换的方法来计算能保证这个精确度的样本的规模。

从某种理论角度看,这种方法没有错误。但是,在实践中,它对打算进行真正研究的大多数研究者来说是没有什么帮助的。首先,通常不会仅仅根据估计值的精确度的需要来确定样本规模。大多数调查研究设计是要得出一些估计值,而这些估计值所需的精确度可能是不一样的。另外,让研究者以相当概括的方式来指定所需的精确度是不常见的。除非是特例,一般情况下不能提前指定一个可接受的误差容限。即使在特例的情况下,所适用的方法也意味着抽样误差是调查评估中唯一的或主要的误差来源。当指定样本调查的精确度时,往往会忽略抽样之外的其他方面的误差来源。这种情况下,单以抽样误差为基础的精确度的计算是不现实的、过度简单化的。而且,在研究资源一定的情况下,减少用于回答率、问题设计或数据收

集质量的资源去增加样本的规模,反而会降低精确度。

与样本规模有关的抽样误差的估计值,在对样本规模究竟应该多大这一问题的分析中起着一定的作用,但是,这个作用是很复杂的。

决定样本规模的首要条件是一个分析方案,这个分析方案的关键部分通常不是有关整体样本的置信区间的估计值,而是总体内那些需要有单独的估计值的子总体的概况,及总体分数在这些子群体的分配比例的估计值。在一般情况下,设计过程很快就进行到总体内那些我们需要有它们的数字的较小的群体,这时,研究者必须估计,为了使这些较小的子群体能有最低限度的精确度,样本究竟应该有多大? 大多数样本规模的确定主要关注的问题并不是整个总体的估计值,而是那些重要的最小子群能容许的最小样本规模。

在确定了最小子群体的最小样本规模之后,接下来我们需要再看一下表 2.5,我们要的是样本量连续统中的低端而不是高端。50 个研究对象够不够? 如果有人研究表 2.5,就可以发现在样本规模在 150 到 200 之间时,其精确度在稳步提高。过了这个点,增加样本规模所获得的效果就小得多。

像许多与研究设计有关的决策一样,对于任何一个给定的研究来讲,我们几乎无法得到一个有关该研究的样本规模究竟应该多大这一问题的肯定回答。有许多方法可以增加调查估计值的信度,增加样本规模就是其中之一。虽然我们不能说出什么是唯一正确的答案,然而我们却可以说出三种决定样本规模的方法是不适宜的。指定一个总体的分数来确定样本规模的方法绝不是确定样本规模的恰当方法。抽样误差从根本上讲,取决于样本规模,而不是取决于总体体现于样本中的比例。说一个特定的样本规模是研究一个总体的通常的或典型的方法,实际上也总是不正确的。在确定样本规模时,制订体现研究目标的分析方案才是关键性的第一步。最后我们认为,用计算某个变量的整个总体的期望的置信区间来确定样本规模的大小方法,同样也是不可取的。

作为总调查误差一部分的抽样误差

抽样过程以三种不同的方式影响到调查估计值的质量:

- 如果抽样框排除了一些要描述的人,那么样本估计值就会由于丢失了这些本应包括在内的人而产生偏倚,偏倚程度取决于他们与那些包括在样本中的人的差异程度。
- 如果抽样过程不是概率的(probabilistic),那么样本和被抽样的对象之间的关系就是有问题的。有人可能用其他理由而不是用抽样过程来证明样本的可信度,但是,除非抽样过程给每个选择的人一个已知的选择概率,否则我们就没有统计上的根据说样本对被抽样的总体而言具有代表性。
- 概率样本的规模和设计,以及被评估内容的分布,共同决定了抽样误差的大小。这就是说,偶然变差之所以发生,是因为我们只搜集了一个总体样本的数据。

抽样误差常常被看作是调查估计值中不可靠性的唯一来源。对于使用大样本的调查来说,产生误差的其他原因可能要更重要一些。本书的主要观点就是非抽样误差应该与抽样误差一样受到重视,我们还经常可以看到这样的情况,在样本设计涉及集群,甚至根本不是什么概率样本时,也能看到有关假设为简单随机样本的抽样误差报告。具有讽刺意味的是,用这样一些方式报告的抽样误差的估计值,会引起读者对样本估计值准确度或精确度的误解。

如果一个详细的名单被用作抽样框,或者采用一个简单的随机或系统抽样方案,或者所有的回答者都是以同样的概率选出来的,那么从一个样本中得出抽样和分析数据可能是十分简单的。就这样的设计而言,表 2.5 和其所依据的公式将会为抽样误差提供一个准确的估计值。但是,甚至用这样简单的设

计，在评估调查估计值的精确度时，研究者仍然要考虑到所有的误差来源，包括抽样框、无回答和回答误差（这些将在下一章讨论）。此外，当对最好的抽样方法有怀疑，或是采用简单随机抽样有偏差时，一位抽样专家的参与，对设计一个合理的抽样蓝图和正确地分析一个复杂的样本设计的后果都是极其必要的。

练 习

为了能更好地理解抽样误差的含义，试从同一份清单（如一本电话簿）中重复抽取规模相同的系统样本（以不同的随机起始点）。将这些样本中有某些特点（如营业性清单）的部分放在一起将形成一种分布。对规模如抽取的样本而言，这个分布的标准差是表 2.5 列出的同容量样本一半。计算表 2.5 中列出的若干条目（如不同的样本规模和比例），对理解这些数目是如何出来的也是很有帮助的。

第3章　无回答：实施样本设计

无法从选入样本中的人收集到数据的比例过高是调查误差的一个主要来源。本章将讨论在邮件、电话和个别访谈中如何与回答者取得联系，并争取得到他们的合作的方法。我们也要对与无回答相关的偏倚问题进行讨论。把这种偏倚看作是那些避免为得到高回答率而做出的努力而采用的一些策略，譬如定额样本的缺陷。

所谓概率样本是指在总体中(或至少在抽样框中)的每个人都有一个已知的被包含在收集的数据中的机会。一个抽样程序会指定一组特定的个体(或某种单元),而样本数据的质量取决于实际上被收集到数据的人在那一组中所占的比例。在决定样本数据对总体描述的质量时,用于收集数据的程序,与样本选择的过程同样重要。

当然,任何一个调查估计值的精确性都取决于什么人对这一特定的问题提供了答案。在每个调查中,有些人虽然愿意成为回答者,但不愿意对每个问题都做出回答。尽管在通常情况下个别问题的无回答率都比较低,但有时也会很高,并且对估计值确实有影响。本章所关注的并不是那些没有回答个别问题的人,而是那些没有提供任何数据的人。

被选作样本但实际上没有提供任何数据的人可分为三类:

- 资料收集过程中没有联系到,因此,没有为他们提供回答问题的机会的人。
- 拒绝提供资料的人。
- 不能按要求提供资料的人(例如:那些病情严重而不能接受访谈的人;使用与研究者不同语言的人;因为在读写技能上有障碍而不能填写自填问卷的人)。

研究者确定使用的程序不仅对样本中实际提供信息的个体的比例(即回答率)大小会有很大影响,而且也会对无回答会在多大程度上将偏倚引入样本数据有影响。值得注意的是,与第 2 章讨论过的抽样误差不同,无回答可能导致样本的偏倚,也就是说,使样本系统地不同于它们所由抽取的总体。本章将讨论无回答对调查估计值的影响和减少无回答的措施。

计算回答率

回答率是评估数据收集工作成败的基本参数。它是接受

访谈(或回答者)的人数除以样本的人数(或单元数)。分母包括了从总体中抽取的所有人,即包括那些虽被抽中但因拒绝、语言问题、疾病或者缺乏易得性而没有回答的人。

有时样本设计会使用某种筛选的方法来从总体中选择研究对象。那些不应包含在研究总体中的筛选单元不能进入回答率的计算。所以,空房子、未运行的电话号码以及没有符合条件的人居住的住户(例如,在对老人进行抽样时没有老人的住户),在计算回答率时都应被排除在外。如果存在某些无法获得用以确定合格的回答人的信息的单元,那么,回答率便会是不确定的。在这种情况下,在计算回答率的时候最好能采用保守,及对未作筛选的单元的合格率做自由假设的方法,并在报告最佳估计值的同时,报告该值的可能范围。

回答率在报告中通常被表达为一个选出的样本中能收集到资料的人所占的百分比。有时我们可以对它做进一步的计算,即计算样本代表的总体分数。如果抽样框没有排除研究总体里的任何人,那么回答率就等于样本代表总体的百分数。但是,对一个电话调查而言,假如总体中只有95%的人有电话,那么,样本代表总体的百分数的最精确评估值应该是0.95乘回答率。

弄清计算回答率的详细方法是十分重要的,因为计算方法的不同会导致比较上的困难。例如,某些机构报告的是"完成率",即在"联系到的"家庭中,访谈完成的比例。由于存在着能抽取却无法联系的单元,因此这个比例总是高于前面提到的回答率。由美国民意研究协会(AAPOR)出版的一本书对回答率的计算和报告方式做了极其详尽的讨论(AAPOR,2000)。

与无回答有关的偏倚

无回答对调查估计值的影响取决于无回答的比例及其导致的偏倚,即它们系统地不同于整个总体的程度。如果大多数

被选到的人都提供了资料,那么即使没有回答的人是很独特的,样本估计值仍然会是非常准确的。例如,美国人口普查局进行的全国卫生调查就成功地完成了将近 95% 的选定的住户的调查。我们不难证明,即使那没有回答的 5% 的人都很独特,得到的样本仍然会与总体非常相似。

人口普查局的经验是一个正面的极端的例子。而邮件调查则是另一个极端的例子,在邮件调查中,有时我们只能从 5% 到 20% 被抽取的人那里得到回答。在这样的情况发生时,最终的样本与最初的抽样过程间几乎没有关系,那些回答人实际上都是自我选择的,而诸如这样的过程是不可能提供任何关于总体特征的可靠的统计数据的。

大多数调查研究项目处于这样的两个极端之间。通常,在农村地区的回答率比在中心城市的高。从一个家庭中有责任心的成年人那里获得资料比对预先选定的特定调查对象进行访谈来得容易。一些主题(例如,健康)可能比其他主题(例如,经济行为或公众意见)能引起更多人的兴趣。另外,不同调查研究机构为提高回答率所投入的时间和金钱的差别很大。

并不存在可接受的最低回答率的统一标准。联邦政府管理和预算办公室评价那些与政府相关的调查活动,通常要求回答率超过 75% 。在美国,学术调查机构通常能在一般住户样本中被选定的成人那里获得 75% 的回答率。市中心区或随机数码拨号调查的回答率可能比较低——通常是低得多。

在邮件、电话和个人访谈中,与无回答有关的偏倚也有所差异。一般的规律是,支持邮件调查的人大多是对调查主题或研究本身特别有兴趣的人,他们比那些没有兴趣的人更有可能寄回问卷。这就意味着低回答率的邮件调查所产生的偏倚可能直接与研究目的有关(例如:Filion,1975;Heberlein & Baumgartner,1978;Jobber,1984)。一项针对那些没有寄回关于健康治疗经历问卷而后来又通过电话进行了补充访谈的人的研究表明,与回答者相比,无回答者通常更年轻、更健康、很少使用保健器材并且大多为男性(Fowler,1998)。

另一个被经常引用的,由邮寄问卷的低回答率导致的显著偏倚的例子是《文摘》(*Literary Digest*)杂志在1936年总统选举中进行的一个民意测验。这个民意测验预测兰登(Alf Landon)将会获胜,而结果却是富兰克林·罗斯福(Franklin Roosevelt)以压倒多数的选票获胜。据说,这些调查的样本是从电话簿里抽取的,而1936年,在家里安装了电话的共和党(兰登所属的党派)成员很可能比民主党成员多。除此之外,《文摘》杂志使用了邮件调查,无回答率也是预测失败的原因之一,只有少数的人按要求寄回了问卷。这点在邮寄问卷的调查中很典型:那些希望处于劣势的候选人获胜的人,即兰登的支持者,特别希望能表达他们自己的观点(Bryson,1976;Converse,1987)。

另外一种在邮件调查中经常出现的偏倚是,受过良好教育的人会比受过较少教育的人更快地寄回问卷。因此,任何与教育(它又与收入水平有关)相关的变量的邮件调查的评估都可能带有偏倚,除非采取措施以便获得高回收率。

与邮件调查相比,调查对象的可得性(availability)对电话调查和当面调查无回答的影响更大。很明显,如果资料收集是在周一到周五的早上9点至下午5点之间进行,那么得以进行面谈的人将具有一些共同的特点:由于那时在家的往往是一些没有工作的人,因此很可能会使调查中的家庭妇女、带孩子的母亲、无业人员和退休人员的比例偏高。这些人一般不忙于义务活动和社会生活,他们更可能是一些很小的孩子的父母。人数较多的大家庭比只有一两个人的小家庭更有可能有人在家。

调查对象在可接近性(accessibility)上的难易程度也会导致与无回答有关的偏倚。使用个人访谈的国家调查,在市中心区获得的回答率总是比在郊区和农村地区的低。导致这种情况产生的原因主要有三点:第一,市中心区难以找到的单身的个体的比率较高;第二,在市中心区,越来越多的个体居住在高层公寓中,使得研究者很难直接与那些人联系;第三,市中心区有很多地方不适合访谈员晚上去访谈,因此,他们很难有机会在那些不容易找到的人在家的时候找到他们。

在过去的 20 年里，市中心区持续增长的人口和其他广泛的社会变迁（单亲家庭越来越多，有孩子的家庭越来越少，女性劳动力越来越多）使得高回答率越来越难达到。有迹象表明，回答率一般都在下降。某些其他机构虽然将回答率保持在原有的水平，但都增加了额外的投入和工作量。格罗夫斯和库珀（Groves & Couper，1998）对无回答的全国性趋势做了完整的总结。

有些材料表明，使用电话调查可以缩小市中心区与农村地区间的回答率差距，因为它有可能比较全面地覆盖城市家庭，能够与居住在保安措施严密的建筑里的人取得联系，并可以通过不懈的努力找到单身户中的人。

另一方面，受教育少的人不是十分愿意接受随机数码拨号的访谈方式，年龄超过 65 岁的老人亦是如此。这种偏倚很少出现在个人访谈调查中。格罗夫斯与凯恩（Groves & Kahn，1979），坎内尔、格罗夫斯、马格利、马斯维茨和米勒（Cannell，Groves，Magilavy，Mathiowetz，& Miller，1987），格罗夫斯与利伯格（Groves & Lyberg，1988）以及和格罗夫斯与库珀（Groves & Couper，1999）对电话调查中的无回答偏倚作出了比较正确的评估。

最后，偏倚还与那些不能接受访谈或不能填写表格的人相关。这些人通常只占一般总体的一小部分。在估计医疗保健的利用情况或医疗费用时，将住院者排除在调查之外这样的丢失可能会造成很严重的后果。在某些特定的邻里地区，将那些不会说英语的人排除在调查之外，也会对调查结果产生明显的影响。如果没有采取专门的措施去从特定的组群中收集资料，样本估计值就会只可应用于一个更受限制的总体：这个总体就是在实施给定的数据收集程序时，实际上有可能回答问题和提供数据的总体。

虽然在我们已经了解由调查员进行的调查中，特别是在基于随机数码拨号的电话调查中回答者与无回答者之间确实存在着人口统计学上的差异，但是我们对无回答究竟会对调查估

计值有什么影响却仍然知之甚少。基特、米勒、科弗特、格罗夫斯和普雷瑟(Keeter, Miller, Kohut, Groves, & Presser, 2000)对两个电话调查结果进行了详细比较:一个调查的回答率为36%,另一个为60%。调查涉及的问题包括了人们对政治和社会的态度等。这两个调查的估计值之间很少有统计上的显著差异。

关键的问题是无回答与调查主题相关的程度。回答者对研究主题的兴趣可能对邮寄问卷调查的回答率起的作用,比调查员获得被调查人的合作起的作用更大。这一点似乎不仅是很有道理的,而且也与现有的证据相一致。总而言之,迄今为止我们还没有掌握可以可靠地预测无回答究竟在什么时候,以及多大程度上影响调查估计值的信息。

表3.1提供了一个例子帮助我们了解无回答对调查结果的潜在影响。假设某样本有100个人,但只有90个做出回答的人(回答率为90%)。在这90人中,有45人对某个问题给出的答案是"是",其他的45人则为"否"。还有10个人(无回答者)的观点我们不知道,假如他们都说"是",那么事实上就有55%的人说"是"。因此,如果:

- 回答率 =90%
- 50%的回答者说"是"

表3.1 由无回答引起的潜在的误差

	回 答*(知道答案)**	无回答(不知道答案)	总的样本(假如每个人都回答时的答案的可能范围)
"是"	45	0~10	45~55
"不是"	45	0~10	45~55
总数	90	10	100

注:*假设回答率是90%

**假设有50%的回答者说"是"

如果在样本中说“是”的人都会回答问题,那么原来的样本中回答“是”的人的实际比例可能在 45% 到 55% 之间。表 3.1 中右边的一列,列出了这一例子中回答“是”的比例的可能范围。

表 3.2 揭示了回答率与范围之间的逻辑关系。我们可以从中看出,当回答率低的时候,由无回答引起的潜在误差就会很大。

表 3.2　在样本中 50% 的人给出某一答案时,不同回答率的可能真百分比的范围

前　提	回应率				
	90%	70%	50%	30%	10%
假设回应者中有 50% 的人做出了某一项回答,再假设样本中每个人也做了应答,那么真值的范围应是:	45% | 55%	35% | 65%	25% | 75%	15% | 85%	5% | 95%

资料来源:根据杰克·埃利森(Jack Elison)和米切尔·D·埃利森(Mitchell D. Elison)所制的有关个人交流表改编。

将无回答偏倚的影响,如表 3.2 列出的那种影响,与前一章谈到的抽样误差的影响(见表 2.5)进行一番比较可能不无益处。虽然一般我们都不知道无回答是怎样偏倚的,但是假设无回答是无偏的,不失为一种明智之举。为保证回答率达到一个合理的水平,以及避免采用可能在回答人和无回答人之间产生系统差异的程序的种种努力,无疑对确保调查估计的精确性是十分重要的。

减少电话或当面访谈调查中的无回答率

就像我们可以说,在其他方面等同的情况下大样本比小样本更可靠一样,我们也可以说具有较低无回答率的调查,可能

比无回答率较高的调查产生一个更好、偏倚更少的样本。无论做什么设计决定，研究者都必须考虑需为减少无回答做出多大努力。

在电话和个人访谈调查中为了得到较高的回答率，必须处理好两个不同的问题：与选定的个人进行联系和取得他们的合作。

为了减少由于无法联系而产生的无回答率，我们可采取如下措施：

- 集中在晚上和周末进行多次拨叫，打电话的次数取决于当时的情况。在城区，每家至少打 6 次。对于电话调查而言，更多次的拨叫的费用可能比较低廉，因此许多机构把 10 次拨叫作为最低限度。而多于 10 次的拨叫也是经常可以看到的。
- 给调查员一个灵活可行的访谈时间表，使他们能在被调查人任何他们感到方便的时候，约他们进行访谈。

而为了取得合作，我们可采取如下措施：

- 如果可能，先寄一封预告函。这可以打消一些回答者的顾虑，访谈员也可因此感到更有信心。
- 提供有关研究项目目的的简要而精确的信息，让回答者明白他们的帮助是十分重要和非常有用的。
- 要确保调查任务和数据的使用不会对回答者产生不利影响。
- 对访谈员进行有效的培训，确信他们已经懂得回答率的重要性。尽快发现那些在回答率方面有问题的访谈员，并对他们进行重新培训，或不再使用那些效率不高的访谈员。

既要坚持不懈地努力，又不能对那些不愿参与的回答人操之过急，在这二者之间保持适当的平衡是不容易的。调查中的参与应是自愿的，而不愿意参与，往往是由于对调查缺乏了解。格罗夫斯和库珀（Groves & Couper，1998）对回答者不愿参与访

谈的原因进行了一些有启发性的分析。

最好是在回答者拒绝之前告知他们调查的目的。大多数调查,从某种特定的角度看,都有某些实际的用途。在有人拒绝回答时,便会减少得到的结果的准确性。我们应该要求访谈员努力做到让被访谈者明白他们需要做什么和为什么要这样做。按规定,我们还应该要求那些最初拒绝接受访谈的回答人重新考虑一下我们的要求,这种做法是合乎情理的。有相当一部分的拒绝回答者,其拒绝的原因是因为我们在他们不方便的时候要对他们进行访谈,而不是他们不愿意接受访谈。在那些最初拒绝回答的人中,在我们以后再次对他们提出要求时,有 1/4 到 1/3 会同意接受访谈。

对回答者来说,访谈过程通常都是一次正面的经历。如果调查研究项目是由负责任的组织进行的,回答者的回答就会得到严格保密。通常,调查研究者会像精神病学家和记者保护他们的信息来源一样遵守保密原则。很多调查研究项目都有合理的动机,这使得大多数人都愿意提供帮助。如果访谈员愿意在回答者方便的时候进行访谈,对大多数回答者而言,便不会有时间上的特别压力。最后,大多数回答者报告说被访谈是一件愉快的事。人们喜欢有机会向一个很好的倾听者谈论自己。

随机数码拨号对取得高回答率来说是一个有特别意义的挑战。在使用名单辅助法选择样本时,就可把信寄到已知地址的家庭。但是,就算这样做了,保证人们接电话回答问题仍然是研究者的一个难题。研究者在电话里开头所说的几句话是成功的关键。由于电话常被用作一些推销和募捐的手段,因此人们听到是陌生人打的电话时都会产生抵触情绪。在此情况下,有时即使再真诚的机构获得的回答率也只有 40% 到 60% ,特别是在调查的主题不能立刻吸引人时回答率就可能更低。获得高回答率的困难在于随机数码拨号作为一种抽样方法的内在局限性。

最后,应该注意,事实表明,在访谈以前付给被访谈者费用会增加个别访谈和电话访谈的回答率。从历史上看,报酬最初

是在要求回答者回答不同寻常的问题的访谈研究中使用的,例如,多次访谈。但是,最近的研究表明,在一次性的访谈里,报酬也会对回答率有所影响(Groves & Couper,1998;Singer,Van Hoewyk,Gebler et al.,1991;Singer,Van Hoewyk & Maher,2000)。

减少邮件调查中的无回答率

减少邮件调查中无回答率的问题稍微有些不同。与回答者取得联系是电话和个人调查中十分重要的一步,而在邮件调查中,只要研究者有一个准确的邮件地址就行了,大多数人迟早都会回家取邮件。然而对于邮件调查来讲,最主要的困难是如何在没有访谈员参与的情况下引导回答者完成回答工作。

写信说服大多数人做某事并不是一种很有效的方法,个别联系比写信更有效。现在有大量关于使邮件联系更有效的设计策略方面的文献。是用彩纸还是白纸印刷调查问卷?怎样才能使信笺抬头给人留下深刻印象?怎样签署落款?应该支付回答者什么报酬?有些研究人员在寄问卷的同时寄现金,而另一些研究人员则承诺一收到寄回的问卷就付报酬?回答者的信件应该用蓝墨水写吗?贴邮票的信封比已付邮资的信封好吗?

总之,能让使用邮件的调查问卷看上去更专业、更人性化或者更吸引人的所有措施都会对回答率产生正面影响。总起来讲,关注这些细节是很值得的。迪尔曼(Dillman,2000)对这些问题做了很好的评述。另外,使调查问卷容易完成可能也是十分重要的。有关的设计细节将在第6章讲述,但这里需要指出以下三点:

- 版面要清晰,以便很容易了解如何作答。
- 问题之间有足够的空白,排列整齐有序,易于阅读。
- 回答问题的方式应该简单易行。除非出于他们自己的选

择,一般不要叫回答者提供书面答案。回答问题的方式可以是在方框中打勾,在数目上画圈,或其他一些简单易行的方式。

迪尔曼(Dillman,2000)再一次为我们提供了大量的有关自填式问卷版面设计的指导。

几个已发表的研究指出,提前向回答者支付若干美元的报酬会提高回答率(Fox, Crask, & Kim, 1988; James & Bolstein, 1990)。几乎所有的调查,即使是对医生进行的调查,都表明支付报酬对提高回答率是有一定好处的(Berry & Kanouse, 1987)。但只有提前付钱才能对回答率有所影响。延迟付款、提议将报酬捐赠给慈善团体和先登记后支取都不能增加回答率(Warriner, Goyder, Jersten, Hohner, & McSpurren, 1996)。最后我们想要强调的是,所谓好的与差的邮件调查,其最大区别在于研究者与无回答者反复进行联系的程度不同,这一点是毋庸置疑的。正如迪尔曼(Dillman,2000)加以概括的、一系列合理的步骤应该包括以下几点:

- 在最初的信件寄出 10 天以后,给所有无回答者寄提醒卡(reminder card),以强调研究和高回答率的重要性。
- 在寄出卡片大约 10 天后,再给剩余的无回答者寄出一封信,并再次强调高回答率的重要性并附上问卷,以防有些人已经将第一次寄去的问卷扔掉了。
- 如果回答率仍不让人满意,也许下一步最好就是给无回答者打电话。如果没有电话号码或者打电话的花费太大,追寄劝说函、发送夜间电报、通宵递送的邮件或其他补救措施都将是有所助益的。

让回答率达到合理水平的困难在于它取决于样本的性质、研究的特点、人们的积极性以及问卷填答的难易程度等诸多因素。如果样本是由有积极性的受过良好教育的个体组成,那么,想要得到较高的回答率便会比较容易。迪尔曼只使用邮件,就在一般总体样本中得到了 70% 的回答率(Dillman, Car-

penter, Christensen, & Brooks, 1974)。当然,如果再给无回答者打电话,回答率将会更高。如果研究者坚持努力,保证研究是合理地精心构思和仔细设计的,邮件调查也可以得到可接受的回答率。

如果研究者打算再次和无回答者取得联系,就必须知道哪些人没有寄回问卷。这个过程不需要拐弯抹角或者复杂化,在问卷或者寄回的信封上写上简单的识别号即可。在实际调查中,最好能在附信中告诉回答人这个数字的用途。

有时,研究者可能想向回答者保证,他们的姓名是不会暴露的,这时有一种可供选择的简便方法,既可保证不泄露回答人的姓名,又使跟踪访问仍可得以进行。给回答者寄去一份没有识别码的问卷,与此同时附上一份单独的有回答者代码的明信片,其内容如下:

> 敬爱的研究员,我在寄回问卷的同时寄回了这张明信片。鉴于我的问卷完全是匿名的,但是这张明信片会告诉你不用给我再寄索要问卷的提醒卡了。

这种方法保证了回答者的匿名性,同时又告诉研究人员某人完成了问卷。有人也许会认为回答者会只寄回这张明信片以避免再收到提醒信,但是这种事情很少发生,事实上,寄回的明信片数量和问卷数量大多数情况下几乎是一样的。希伯(Sieber, 1992)探讨了各种各样的用以匿名地收集数据的方法。

减少互联网调查中的无回答率

尽管互联网调查是一个相对较新的方法,没有像邮件和访谈调查那样的经验性的文章,但是其难度和挑战可能接近于邮件调查和访谈调查。问题仍在于要在没有访谈员的情况下促使人们做出回答。有积极性和对研究感兴趣的人最有可能做出回答。使填答简单易行、反复进行联系、使用多种联系方式

以及为那些最初没有回答的人提供可选择的回答方式,这些可能都是最大限度地提高回答率的关键。像邮寄问卷调查一样,调查内容与被调查总体间的一致性以及调查样本在互联网使用者中的易得程度,可能都是保证获得高回答率的重要因素。迪尔曼(Dillman, 2000)对这些问题做了详细的概括。

多方式数据收集

要最大限度地减少调查无回答的最好方法之一就是使用多种方式相结合的方法收集数据。诚如我们已经知道的那样,多方式数据收集的关键也是联系调查对象的途径、被调查人的积极性和数据收集的费用。多种方式结合的方法可以让研究者联系到那些用一种方式很难找到的人,同时也能从本来没有多大内在积极性的样本成员那里收集到资料。例如,一个吸引人的方案是使用电子邮件或邮寄问卷作为数据收集的第一步,接着对无回答者进行电话访谈。另一种很有效的设计,是把电话和面对面的访谈结合起来。

多方式调查中的一个关键问题是用各种方式得到的数据,彼此要具有可比性。有的问题的答案会受到数据收集方式的影响,有的则不会。为了将通过不同方式收集到的数据整合起来,数据要具有可比性这一点是十分重要的。在第 4 章,我们将就这些问题进行更多的探讨。

修正无回答率

有些无回答是不可避免的,这种情况下,有三种尽可能将无回答导致的误差降低的方法:使用代理回答者(proxy respondent)、进行统计调整,和对无回答者重新进行调查。

代理回答者

很多调查一般是从某个家庭的回答者那里收集其他家庭成员的资料。如果一位回答者不能或不愿接受访谈，一种选择就是让该家庭的另一个成员来代替那人回答。对代理回答得到的资料的质量的大多数研究表明，对于大多数话题，代理回答不如回答者自己回答的效果好。此外，只有很少的研究者会接受代理回答者对主体态度诸如感受、知识或意见的报告。但是，对于事实性的资料，如果有一个现成的知情的代理者，那么使用代理回答者也能够有效减少无回答导致的误差。格罗夫斯(Groves,1989)对那些有关代理回答人的尚无定论的文章做了较为全面的评论。

统计调整

由无回答所引起的统计调整工作如下：假设某总体中有20%的成人是65岁或65岁以上的老人，由于各种无回答的存在，样本中却只有10%的回答者在65岁或65岁以上。在调查完成以后，研究者可以给那些在65岁或65岁以上的回答者的答案加权，以便使他们与20%的回答者等效。鉴于65岁或65岁以上的人的答案往往会与年轻人的答案有所不同，经过这样的调整的最终估计值，可能会比未作调整的好一些。但是，这个结果需要假定65岁或65岁以上的回答者和该年龄段的无回答者是相似的，而这个假定有可能不是真的。

很多机构是通过统计处理来降低由无回答引起的估计值的误差的。卡滕(Kalton, 1983)、利特尔和鲁宾(Little & Rubin,1987)、格罗夫斯(Groves ,1989)仔细探讨了这种调整的技巧、设想和局限性。毫无疑问，这种调整会提高某些估计值的精确度。即使如此，大多数人都会认为，减小调查估计值中由无回答导致的误差的可靠方法的首选，还是降低无回答率。

调查无回答者

假设在一次邮寄问卷调查中,被抽样的人中有60%做了回答。研究者认为如果对无回答者进行电话或个人访谈,其中很多人就有可能回答问题,但苦于缺乏足够的资金对所有的无回答者试一下这些方法。这时研究者可以退而求其次,抽取一个无回答者的样本试着用一些费用较高的方法与样本中的个体取得联系。

对无回答者进行调查的工作有两种不同的用途:第一,第二轮收集到的数据可以用来评估最初样本里的偏倚的方向和数量。当然,第二轮数据收集也会存在无回答问题,这使得来自第二轮数据收集的数据不能完全代表所有的无回答者。但是,尽管有这样的局限,这些资料还是可用于改进在这里讨论的那种统计调整。第二,如果第二轮的数据收集中所用的问题重复初次调查中的问题,那么它的数据就可以补充到最初始样本的数据集中去。为了将这些数据追加到初始样本的数据中,我们必须对它加权,以对只有一个无回答者样本做了这样的处理事实作出调整。

如果对一半的无回答者做了补救调查,那么在将它们并入与最初收集的数据时,应对在这一阶段中得到的数据用因子 2 为权进行加权。此外,我们可以把这个结果看作一个调整的样本总体百分比(一个调整后的回答率),计算公式如下:

$$\text{调整后的回答率} = \frac{\text{最初的回答数} + 2 \times \text{来自无回答样本的回答数}}{\text{最初符合条件的总样本数}}$$

非概率(或修正的概率)样本

前面各章的讨论都基于概率样本设计这一假设,因而回答者都是通过某种客观程序选定的。研究者要解决的问题是收

集那些被选定的对象的资料。但是,也有一种普遍使用的选择方法(或一系列方法)是不选定特定个体的,因此也就无法计算回答率。我们之所以要在这里讨论非概率样本,主要是因为除了回答率不可计算外,它们与那些回答率很低的样本大致相当。而无法计算回答率,可能导致使用者不知道他们使用的资料的局限是什么。

调查研究者之间关于概率抽样的重要性的争论是很有趣的。联邦政府通常不会为那些不是建立在概率抽样技术基础上的对总体特点进行评估的调查研究提供经费。大多数学术性调查机构和很多非赢利性调查机构都有类似的抽样方式。同时,大多数较大的民意测验团体、政治民意测验团体和市场研究机构主要使用非概率抽样方法(参见 Converse,1987,对这种区别的历史根源进行了探讨)。

概率抽样的基本原理是入选样本是以预定的程序为依据的,这个程序为定义的总体成员设定了一个选择率。此外,在概率抽样中,回答者的特点或研究人员的处置方式都不会对某人入选样本的可能性有影响。虽然,非概率抽样的修正方式不完全相同,但是,在最后阶段,都有一个共同的特点,那就是原本不是样本设计的一部分的调查员的处置方式,和(或)被调查人本身的某些特点有可能会影响入选样本的可能性。下面我们将对两种比较常用的非概率抽样方法进行讨论。

对涉及非概率抽样的个人访谈研究而言,一个研究者可能以一种与一个抽样者在为一个区域概率样本抽取街区时使用的十分相似的方法抽取街区。其区别在于,一旦某个街区被抽中,访谈员将被派访问那个街区,并在居住在那个街区的住户中完成某一既定数目的访谈。在没有那个街区的具体的住户名单时,有一种可能采取的办法是给访谈员访问位于该街区的任何住户的自由,且不要求调查员在访问未遇时再次进行回访。

在电话调查中也有可能使用与之相似的方法。选择一个特定的电话局,或者一个电话局里的一组号码,其目的都是为

了完成一定数量的访谈。如果在电话拨通后没有人接电话,或电话接通时没有可接受访谈的人时,就拨叫同一组中的另一个号码,如此这般,一直进行下去,直到得到要求的访谈数足够为止。在抽样的第一阶段,如果像前面所指出的那样,将样本的分布在某种程度上,以某种方式围绕总体的地理区域分布,那么在选取住户和回答者时,可能会引进三种十分明显的偏倚。

在个人访谈而不是电话访谈中,访谈员可以自由选择去哪家访谈。事实证明,访谈员更喜欢去那些富有吸引力的住户,更愿意去那些住在一楼的家庭,而不是二楼或者三楼的,愿意去那些不养狗的家庭。影响访谈员选择访谈对象的其他一些因素,留待读者自己去想象。

一些研究机构试图通过指定从哪里开始以及规定不得略过哪些家庭单元的办法,来限制访谈员的处置权。但是,由于预先没有该街区抽样单元的名单,事实上不可能判断访谈员是否执行了这些指令。另外,如果想预先有该街区抽样单元的名单,那些本来用这种方法可以节约的成本就会被抵消掉。

非概率方法的第二种偏倚性是对可得性的影响。假如访谈员不对那些先前没人在家的家庭进行回访,那么在访问时平时有人在家的家庭被选中的机会就比那些平时没人在家的家庭高。性别、年龄、工作状况、家庭规模和有否孩子等问题都与可得性有关。

未加控制的抽样会产生明显的样本偏倚。提高样本质量的最常用的方法是对明显偏倚引入配额。例如,要求访谈员对特定街区或电话号码组中的一半男性和一半女性进行访谈。偶尔,也会设定一些额外的限制条件,例如预期的种族构成、成年人中老人和年轻人的数量。重要的是不给配额制定太多的强制条件,否则访谈员会用很多时间打电话或在街区里徘徊以寻找符合条件的回答者。

配额或非概率样本的最后一个固有偏倚与取得合作有关。如果回答者说自己很忙或不方便接受访谈时,就说明访谈员没有设法激起对方合作的意愿。如果没有对项目进行有效的介

绍,总体中的相当一部分人不会对提供帮助感兴趣。没有努力向人们介绍的研究项目很容易导致他们拒绝合作,这种情况下,由于缺乏那些忙于工作的人、对研究或对研究的特定主题不了解或不感兴趣的人,都会导致样本偏倚。

萨德曼(Sudman,1967,1976)认为,包括那些努力与无回答者进行联系的调查在内,所有的调查中都会存在无回答的情况。一旦知道某个人不愿意合作或打了几个电话都无法联系到,萨德曼建议从同样街区再找一个代替者,这事实上会提高估计值的质量。他认为用样本中的邻居做替代比指定一个人或找不是样本中的邻居做替代更好。当对访谈员的处置权施行控制时,如萨德曼所说,减少回访所节省的开支中的相当一部分被管理费用的增加所抵消了。

非概率抽样方法可以为个人访谈调查节约成本(对电话调查而言,节约的开支较少)。另外,就可比的那些方面看,其得到的样本数据与概率样本得到的相仿。即使如此,有两个事实值得注意:第一,由于节约成本的关键在于避免回访,由此导致在大多数非概率样本民意测验中,总体中只有1/3的人有机会被抽中(即第一次电话联系就在家的人)。一个只给总体中1/3的个体被抽取机会的样本,也就是具有已知不同特点的1/3的总体,很有可能在很多方面并不具有代表性,因而会对样本统计值差产生影响。鲁宾逊(Robinson,1989)提供了一个被认为做得很好的非概率样本却歪曲了事实的例子。他比较了两个调查的结果,一个调查是采用概率样本,另一个是采用非概率样本,它们的目标都是对艺术的兴趣和出席与艺术有关的活动的评估。非概率样本调查大大高估了人们对艺术的兴趣。第二,概率理论和抽样误差的假设常被吹捧为可以描述非概率样本的信度,其实它们对非概率样本并不适用。如果有代替者存在,样本就不是一个概率样本,尽管它有可能以合理的实际方式分布于总体中。

有时非概率样本也是有用的。亨利(Henry,1990)描述了各种各样的非概率样本以及它们的适用条件。然而,如果研究

者决定使用非概率样本,他应该告诉读者样本是怎样被抽取的,样本在易得性上和自发接受采访的过程中可能产生偏倚,以及计算抽样误差的一般假设在这个样本中的不适用性。这样对读者的提醒目前还不普遍。在很多情况下,非概率样本完全不具有代表性,因此对社会科学研究的信度构成了严重威胁。

无回答是误差来源之一

无回答是一个复杂而重要的调查误差的来源。表 3.2 表明了无回答对研究结果的重大影响。虽然我们可以计算回答率,却不知道无回答对数据的影响程度。基特(Keeter, 2000)等的研究表明,一个比较低回答率的调查结果与一个回答率高得多的调查结果是非常相似的。

关键问题是我们常缺乏关于无回答何时可能或不可能产生关于调查内容偏倚的准确资料。由于缺乏这些资料,所以有关最大限度地提高回答率的争论主要集中在信度问题上。回答率高的时候,由无回答引起的潜在误差的可能性就小;回答率低的时候,无回答就有可能产生重大的误差,于是调查结果的批评者就有了有力的证据说数据是不可信的。

练　习

如果把一个住户样本的选择作为抽取 18 岁以上成年人样本的第一步,回答率是完成访谈的人数除以样本中被抽样程序选定的参与研究的个体的总数。你在计算回答率时会把以下各组包括在分母里还是排除在分母外?为什么?

- 无人居住的家庭
- 出去度假的人
- 临时住院的人

- 拒绝访谈的人
- 居住者都小于18岁的住户单元
- 操与研究者不同语言的人
- 家庭其他成员说他有精神病或者神志不清不能参与访谈的人
- 在访谈员打电话时从都不在家的人
- 离家去上大学的人

定义你的样本统计量(和你的回答率)适用的总体。

第4章 数据收集的方法

数据收集方式——邮寄问卷、互联网、电话、个人访谈还是团体填答的选择与抽样框、研究主题、样本特点、可利用的工作人员和设备有直接关系;它与回答率、问题的形式和调查成本也有关系。电脑可以用于所有的数据收集方式。本章将就数据收集方式的选择涉及的各种问题,及选择某种数据收集方式导致的后果进行讨论。

一个研究者必须作出的、影响面最广的决策是数据收集方法的选择。是让访谈员进行提问和记录，还是让调查对象自行填写问卷？如果在数据收集的过程中使用了访谈员，那么就需要进一步考虑，访谈员是该亲自去还是通过电话访谈。如果回答者在没有访谈员的情况下自己阅读和回答问题，那么，就存在怎样向其提供问卷的问题。有时调查问卷是以集体或个人的方式发放并当场回收的。在住户调查中，以分发或邮寄的方式发放问卷，并以同样的方式回收。如果调查是在互联网上进行的，就可以通过电子邮件或者要求回答者进入某个网站回答问题的方式进行。

尽管大多调查只使用一种资料收集方法，但是，使用多种方法结合的方式也是十分普遍的。例如，有时在个人访谈中，有些问题需要回答者自写，或直接把答案输入便携式电脑。计算机化的数据收集方式已经与电话调查结合起来。为了减少无回答率，对没有寄回邮件问卷的那些人，则由访谈员打电话或亲自登门与其联系；对于那些没有电子邮件地址或没有回答的人就可以用邮件的方式作为补充；对于那些访谈员无法在其家中找到或者已经搬家的调查对象，则可以通过电话或自填问卷的方式进行调查；在一些住户调查中，对可以得到电话号码的家庭可以通过电话访谈，而对找不到电话号码的家庭则使用个人访谈。

各种数据收集的方法都可以与电脑相结合。在进行电话调查和亲自上门调查时，访谈员一般都会用到电脑。更有意思的变化是让回答者把回答直接输入电脑。网上调查是完全没有访谈员介入的。还可以通过让回答者在电脑上回答问题的方式，在其所在的地方——诸如医生办公室等——收集数据。最后，还可以通过电话的电脑语音提示系统提出问题，并由电话的数字按钮来选择答案。

在特定的条件下，每种收集数据的方法都可能成为最好的方法。这一章的目的就是讨论如何针对不同情况选择恰当的数据收集方式。

数据收集策略中的主要问题

抽　样

研究者计划获取样本的途径是与最好的数据收集方法相联系的。抽样方法不同,这样或那样的数据收集方法的实施难易程度也不同。如果我们正在用一张清单进行抽样,那么清单上有什么样的信息这一问题就十分重要。如果名单上没有详细的通讯地址、准确的电子邮件地址或正确的电话号码,那么用与其相应的数据收集方式收集数据就会十分困难了。给每个有电话的住户都有被选中的机会的随机数码拨号法,使电话收集数据的方法的潜力有了很大的提高。假如有人本来就不希望包括那些没有电话的住户,那么随机数码拨号法也许就是一种成本最小的抽取一般住户样本的方法了。

当然,也可以先只用随机数码拨号法进行抽样,并与住户进行最初的联系,然后再通过某种其他方式收集数据。在与某个住户取得联系之后,研究者就应该询问他们的地址,并征得他们的同意,或寄去一份问卷,或派访谈员前往。这种设计十分适合于寻找一个难得的总体的情况,因为对同样的调查任务而言,无论是进行抽样还是筛选户内合格的被调查人,电话调查都比个人访谈经济。诸如这样的数据收集策略的困难之处在于数据在收集过程中的合作率。

在样本的根据是一组地址时,无论这组地址来自一份清单,还是一个区域概率样本,电话、个人访谈和寄邮件的方式都是可行的。当然,如果有详细的地址,就可以采用派访谈员的方式。另外,我们也可以通过某些电话或互联网服务商,得到许多地址。因为公布的电话号码或用户名录大全中,都列有对应的用户姓名和详细住址。

当然诸如这样的途径不可能提供所有地址的电话号码。

因此,就必须用其他的数据收集方式对其加以补充。然而,在很多时候都可以通过电话完成大多数的数据收集,由此不难看出电话调查所具有的潜在优势。如果我们通过一份详细的地址清单进行抽样,那么邮件式调查也是可行的。不过在一所建筑内有多个住宅单位的城市地区,单元号(或住户姓名)及街道地址都是很重要的。没有单元号或住户姓名,寄往多单元结构的住宅的邮件将无法投递,或寄不到正确的单元。现在在农村中也发生了类似情况,大多数农村住户都有了具体的地址,因而一个邮寄式调查是否可行主要取决于我们掌握的有关信息,是否已经具体到足以使邮件能到达选出的特定的单位。

包含了电子邮箱的样本名单,则为数据的收集开通了另一条通道。虽然对于一般总体调查来说,使用电子邮件的方法并不合适,但在很多总体(雇员、学生、专业机构的成员)中电子邮箱不仅十分普遍,而且很容易得到。在这些情况中,把网络作为主要的,或至少作为其中一种数据收集方式不失为一种明智之举。

最后一个要讨论的与抽样相关的问题是明确回答者。如果抽样框是一份个人清单,那么,任何收集方法,包括邮寄都是可行的。但是有许多调查都需要在数据收集的时候明确被调查人。如果把问卷寄向一个住户或组织,那么研究者本人就很难控制问卷的实际完成人,因此,如果对调查来讲,明确被调查人是一个重要的问题,那么调查员的介入将会对这一问题的解决有很大帮助。

总体的类型

总体的读写能力和合作的动机是选择数据收集方式时要考虑的两个突出的问题。相对于利用访谈员的方法而言,自填问卷就要对回答者的读写能力提出一定的要求,如果考虑利用网络收集资料,那么,对回答者而言,除了读写能力之外,还需要掌握一些电脑的使用技术。回答者中那些没有受过多少教育,英语的读写水平不高(但会说英语),视力不好,不经常使用

电脑，以及易于生病或劳累的人都会认为使用访谈员比自填问卷的方法要更简单易行。

就使用自填问卷的方法而言，存在的另一个问题是问卷回收。如果没有访谈员的介入，回答者的内在动机就对返回问卷的行为起决定性作用。对某研究问题特别有兴趣的人就更有可能寄回问卷（Fowler，1998； Heberlein Baumgartnen，1978；Jobber，1984）。在这种情况下，如果研究人员是从一群读写能力很强，并且对研究很感兴趣的人中收集数据，使用邮寄或者电子邮件的方法就会更吸引人们的注意；相反，如果是从一群读写能力差或对研究的兴趣和动机都很低的人群中收集数据，那么访谈员介入的方式就更为恰当。

另一个要考虑的问题是要便于联系，访谈员介入的方法的使用取决于访谈员能否与回答者取得联系，并对数据收集作出安排。自填问卷的一个突出的优点就是只要提供的联系地址是正确的，回答者就会收到问卷，而且，即使回答者很忙，他也会在方便的时候完成问卷。假如调查是与工作相关的，忙于工作的人习惯于以电话的方式进行，并且要预约；通常会有人为他们安排时间表。以预约的方式进行电话访谈是从这些工作忙碌的人那里收集数据的最好方法。

问题的形式

一般说来，如果我们打算使用自填式问卷，那么我们就必须使用封闭式问题，也就是在回答问题时，只需要在方框中打勾，或在由研究者预先提供的一系列答案的适当地方上画圈就可以了。从某种程度上说，问题的难易程度会对回答率有很大影响。其次，自填问卷中的开放式答案常常会产生一些没有什么用处的数据。由于没有访谈员参与，因而在遇有不完整的答案时，便无法做进一步的探究，以使答案变得明确并符合问题的要求，因此，回答者之间的答案也没有可比性。不仅如此，这样的答案也很难进行编码。如果说它们最终还有什么用处的话，那也只能作为描述性的材料而无法用作量化的数据。

尽管开放性的问题通常需要有访谈员参与，但在有些情况下仍然可以较好地将其处理成封闭型自填问卷的形式来进行。一个很好的例证是，在研究者要用相似的形式调查大量问题时，如果让访谈员去读一长串相似的问题会是一件十分吃力而乏味的事情。在这种情况下，就应该把那些问题通过问卷或电脑以自填的形式进行。诸如这样的策略将会使访谈变得易于操作。

在问题的答案的类别很多或很复杂时，自填问卷法同样有自己的优势。在个人访谈中，一般要给回答者一张列出了答案的卡片，以帮助他们记住所有的可能答案。而电话访谈则需要作出一定的调整，通常我们用下面三种方法进行调整：

第一，研究者也许只须对电话调查的答案的类别数加以一定的限制。有人认为，在电话调查中最多 4 个类别比较合适，在多数电话调查中，答案以 2 或 3 类为主。

第二，如果电话访谈员慢慢地、一项一项地读出问题的答案，读完一遍之后接着再重复一遍，那么就可以使用一个较长的问题清单，在访谈员宣读的过程中，回答者就可以随即选出自己认为正确的选项。然而，以这种方式获得的答案是否与用可视清单获得的答案一致，目前尚未得到证实。就某些种类的问题而言，答案项的宣读顺序会影响到回答者的回答（Bishop, Hippler, Schwartz, & Strack, 1998; Schuman & Presser, 1981）。

第三，研究者可以将一个复杂的问题分成两个或多个更简单的问题。例如，通常需要人们在超过 4 种或 5 种选项中报告他们的收入情况，对一个有 9 种选项的问题，就可以分成如下两段进行：请问你们全家的收入是少于 30 000 美元呢，还是在 30 000 美元到 60 000 美元之间或是在 60 000 美元以上？然后，根据回答的情况，访谈员要另外问一个有三种选项的问题，诸如：那么，请问你们全家的收入是少于 40 000 美元呢，还是在 40 000 到 50 000 美元之间或是在 50 000 美元之上？诸如这样的问题在编排上的变化的确可以使问题的回答变得容易，但是，有时问题形式本身会对答案有影响（Groves, 1989）。

对那些包括复杂情境或事件的描述和那些需要图片或其他视觉提示的问题形式,就不适合采用电话调查。如果这种量度的确定是调查的重要组成部分,那么就需要使用其他的调查方式,而不是电话调查方法。尽管研究者们已经证明,他们能够使大多数问题适用于电话调查。如果一个调查工具既要用于访谈方式,也要用于自填方式,那么先按照访谈方式设计是明智的,因为在通常情况下,将一个应用于访谈方式的工具应用于自填方式要做的修改,比将一个应用于自填方式的工具应用于访谈方式要少得多。

问题的内容

很多研究把不同资料收集方式所得到的结果进行了比较(例如,Cannell et al.,1987;Groves & Kahn,1979;Hochstim,1967;Mangione,Hingson,& Barret,1982)。德·路乌与范·德·邹文(De Leeuw & Van der Zouwen,1998)以及迪尔曼(Dillman,2000)对结果进行了细致的总结。对于大多数已被研究过的调查问题来说,由个人访谈、电话访谈和自填问卷得到的总计分布是十分相似的。

研究者有依据认为,在处理一些敏感话题时,使用这种或那种方法会更有优越性。自填问卷法被认为是其中最好的方法,因为回答者不必直接向访谈者承认自己反社会或主流价值观的性格或行为。另外一些人认为,电话访谈具有非人格的色彩,这将有助于人们报告一些负面的事件或行为。另外,随机数码拨号的调查过程几乎完全是匿名的,因为研究者没有必要知道回答者的姓名和家庭住址。还有一些人则认为,个人访谈是询问敏感性问题的最好办法,因为访谈员有机会和回答者建立一种融洽和谐的关系及相互的信任感,这些都是回答者回答敏感性问题所需的。

尽管以上所有说法看起来似乎都有道理,但数据却清楚地表明自填问卷法与访谈法相比,前者不仅更多地用于敏感问题的调查,而且数据也更为准确。与有访谈员参与的方式相比,

自填问卷法无论是使用笔答还是电脑辅助，都会产生更准确的结论（Acquilino，1994；Dillman & Tarnai，1991；Tourangeau & Smith，1998；Turner et al.，1998）。并且，对十分敏感的内容（例如，毒品使用和性行为）以及涉及个人形象的更微妙的问题（诸如健康问题或在前列腺手术后的“普遍问题”）的调查中，也显示出同样的效果（Fowler，Roman，& Di，1998；McHorney，Kosinki，& Ware，1994）。如果敏感问题是调查关注的焦点，那么运用一种没有访谈员参与的方法来获取答案肯定会提高估计值的精确度。

若从这一方面来比较电话和面对面的访谈时，数据给出的结论就不那么清楚了（De Leeuw & Van der Zower，1988）。然而，较之面对面访谈，电话访谈更为普遍的用于揭示一种不同于社会意愿的偏向。曼吉温（Mangione，1982）等人发现，当调查人们过去有酗酒问题的比例时，用这两种方式得到的结果有明显差异。霍奇斯汀（Hochstim，1967），亨森、罗斯、坎内尔（Henson，Roth，& Cannell，1977），艾奎利诺（Acquilino，1994），福勒、罗曼和迪（Fowler，Roman，& Di，1998）也发现了相同的结论。

问题内容影响收集数据的方式的另一个与之完全不同的方面是问题回答的难度。在一些调查中，研究者可能会问到一些很难准确回答的事件或行为，原因是时间过去了很久或是事情太琐碎。这种情况下，报告的准确度可以通过查询记录或与家庭中的其他成员进行讨论而得到提高。标准化访谈是一个快速的问答过程，它很少提供这些机会，尤其是在电话访谈中。自填问卷法为调查对象提供了更多的考虑、查找记录以及与家庭成员商讨的时间。

当需要了解关于某些小事件的细节性问题时，例如人们吃什么东西、怎样花钱、看了什么电视节目等，最好的办法就是叫回答者记日记。虽然这要受到记日记人数的限制，而且研究者也会给其一定的报酬，但是这不失为获得细节的好方法。有一种代替书面日记的方法就是使用电脑辅助，其中包括让人们拨

打 800 电话或在网上提交，同样能获得高质量的数据。

总的来说，当样本具有可比性时，研究者发现很多的调查估计值不会受到数据收集方式的影响。除非刚才提到的那些问题在研究项目中特别突出，否则怎么样收集数据这一问题的确定要考虑多方面的因素，而不只是考虑问卷的主题与资料收集方法之间的相互关系。不过，注意问题的形式或内容与数据收集方式之间的相互关系对调查结果的影响，仍然是调查设计过程的一个重要组成部分。

回答率

在数据收集方式的选择中，对回答率的考虑比其他因素更为重要。显然，当使用团体填答调查法时其巨大的优越性之一就是高回收率。一般说来，要求在教室的学生或在工作环境中的工作人员填写问卷时，回收率几乎是 100%。对回收率的影响通常是由旷工旷课或按工作日程表工作班次轮空或休假造成的。

邮件或电子邮件调查的回答率主要取决于研究的总体和调查的目的。在由美国老年保健医疗机构进行的关于前列腺手术病人的调查中，其邮件回答率高达 82%。在给无回答者打电话后，其回答率超过了 90%（Fowler, Roman, & Di, 1998）。当然，这类邮件调查中也有回答率少于 20% 的情况。

毫无疑问，无回答的情况是与使用邮件调查方法有极大关系的。在前一章中提到过，如果研究者只简单地给选取的调查对象寄出问卷却没有再寄一封信或发一份问卷，其回答率大多低于 50%（Heberlein & Baumgartner, 1978）。如果采用了广泛而合理的补救措施，并且对调查进行了精心设计及认真执行，邮件调查的回答率就会与其他方式获得的回答率相似（例如，Dillman, 2000）。

电话调查回答率的高低在一定程度上取决于抽样的方案。有一种电话调查方法就是个人访谈方法的一种拷贝。如果研究者既有回答者的电话号码，又有他们的通讯地址，就可以事

先寄一封信给他们,介绍研究项目和解释研究目的。然后访谈员可以给调查对象打电话以请求合作。在这种情况下,电话和个人访谈的回答率不会有显著的差异。尤其在访谈员能亲自登门造访无回答者时,情况更是如此(Groves,1989;Hochstim,1967;Mangione et al. ,1982)。

然而这里提到的方法只代表了一小部分的电话调查。更多的电话调查是与随机数码拨号抽样相联系的。随机拨号的一个独特之处在于:因为不知道地址,不可能预先通知回答者。即使是有能力并尽责的调查机构所进行的电话调查,其回答率也比通过个人访谈获得的回答率低5%以上(例如:Groves & Kahn,1979;Groves & Lyberg,1988)。

在一些城市地区,因为使用了电话而使回答率有所提高。然而在郊区和农村地区使用电话获得的回答率比使用个人访谈员的回答率低。在有保安系统的楼房里,和访谈员晚上不愿意去的邻里地区使用电话调查可以增强调查的覆盖面。在电话调查中,也可以打10个甚至是15个回访电话去找那个极少在家的回答者。另一方面,正如在第3章里提到的那样,在随机数码拨号调查中有些无回答的情况是与其教育状况和年龄有关的,这种情况在个人访谈调查中则出现得相对较少。

总之,出于对回答者的尊重,如果先寄信告知情况,电话调查方法和个人访谈的方法之间的区别就很小了。另外,当研究者想对先前访谈过的人进行回访以进一步获得资料时,利用电话与通过个人访谈的回答率没有什么差别。在一些人口稠密的市中心区,电话调查方式,即使是采用随机数码拨号,获得的回答率与个人访谈几乎一样(甚至会更好)。

就更广泛的总体而言,随机数码拨号似乎要付出一种代价,那就是选出的住户的回答率将低于个人访谈得到的回答率。在回答率下降5%到20%时,再乘以那些被排除在样本之外的没有电话的住户的比例时,这两种调查法在回答率上的差别就不是微不足道了。研究者在选择随机数码拨号方法时,必须有准备接受它的这一不足,否则必须付出加倍的努力来避免

这一不足。

成　本

邮寄和电话调查的诱人之处在于它每一回报的成本低于个人访谈。当然,成本最低的调查是使用互联网。调查成本取决于多种因素,一些较突出的因素是设计问卷所需的专门时间、电脑辅助程序的编制和调试时间、问卷长度、样本的地理分布状况、样本的可得性和兴趣度、回访的措施、选择回答者的原则以及训练有素的工作人员的可得性。

人们可能会误认为邮件调查费用低,其实,邮资、工作人员邮寄问卷时间的成本以及印刷问卷的成本是一个不小的数目。另外,如果要进行电话提醒,成本将更大。

成本对比中另一个关键是电话费用。使用电话的成本也会影响到个人访谈调查与电话调查二者成本的比较。但是,在同一样本的调查中,多数情况下,个人入户访谈调查比电话调查的成本高。访谈员走访住户以及联系调查人的工资和花费都必然会远远超过电话调查。

邮件调查和电话调查的成本对比也取决于总体状况。对于一个积极性十分高、能迅速地对调查作出反馈的样本来说,邮件调查成本就会低于电话调查的成本。然而,如果就更为一般的情况而言,达到大致相同的回答率,邮件调查或电话调查的费用基本相同。

尽管在邮件调查和电话调查之间进行选择时,诸多必须考虑的因素中并不包括成本,但选择个人访谈调查方式时,成本问题是至关重要的。不过,在许多情况中,个人访谈所具有的优点,使其成为达到一系列既定研究目标的最好方法。

最后,如果调查可以在互联网上进行,那么其成本将是最低的。有些类型的数据收集,最初可能需要对软件进行一些投资。除此以外的主要费用是设计和测试调查工具的工作人员的工时费。当然,在进行成本比较之后,问题的关键便是是否可以通过电子邮件联系到合适的样本,并达到要求的回答率。

工作人员的能力和经验

在选择数据收集方法时必须要考虑工具和工作人员的易得性。培训一个访谈员费时费力，新访谈员培训耗损率一般较高。很多新的访谈员不善于赢得回答者的帮助，从而在刚开始的时候会有很高的拒访率。另外，善于培训和指导访谈员的人也不容易找到。因此，对于那些想进行有访谈员参加的调查者来说，他们必须实际考虑的一个问题就是访谈员收集数据的专业能力。如果有人曾经进行过贯穿整个过程的调查或者有指导和培训访谈员的经验，那么用访谈员的调查就较可行。否则，选用自填问卷法将更有利。

数据收集花费的时间

数据收集的时间是因数据收集的方法而异的。邮件调查通常要 2 个月才能完成。其常规程序包括邮寄问卷、等待、再多寄一些问卷、再等待，最后是打电话或派人上门进行进一步的访谈。当然，网上调查不考虑等待邮件邮寄的时间，但是它通常需要反复进行联系和提醒。极而言之，最快的电话调查可以在几日内完成，而这种最快的电话调查是以很多的无回答为代价的，因为有些人很可能在短期内找不到。然而，规范的电话调查通常都比样本容量大致相当的邮件或个人入户访谈调查要快。

个人入户访谈调查的时间长度不能一概而论，因为它取决于样本范围和工作人员的易得性。但是可以肯定的说，只有在十分特殊的情况下，个人访谈的数据收集的时间才短于电话调查。

电脑辅助的数据收集

传统的调查是依靠访谈员或回答者手中的纸和笔进行的。然而，在过去的 10 年里，电脑取代了纸和笔：调查的问题都被输入电脑，访谈员或回答者通过屏幕阅读，答案通过键盘输入

电脑。电脑辅助资料收集的主要优点在于，答案当即就被储存为机读形式。相对其他调查而言，这种方法有许多优点：

- 电脑可以采用复杂的问题跳越的方式，这是纸笔形式很难办到的。
- 从前面的问题甚至从前面的访谈中得到的资料，可以在后面的问题的文字表述中，或后面列出的一组问题中考虑。
- 电脑能识别出给出的数据中的不一致，并在数据收集时就使它变得一致。

其缺点在于，需要费相当长的时间进行程序设计和测试，才能确保电脑辅助数据收集没有错误，并且如第 8 章将更为详细地谈到的那样，研究者自己没有能力对资料输入过程的质量进行检查和控制。因此，就像大多数关于调查设计的决策一样，电脑辅助资料收集的价值是因不同项目的要求而异的。萨瑞斯（Saris，1991）详细地描述了电脑辅助访谈的成本和优势。库珀（Couper，1998）等展示了大量的与电脑辅助数据收集有关的最新研究。

在进行电话调查时，用电脑作为辅助手段是十分普遍的。事实上，对一些人而言，电脑辅助的电话调查（CATI）从本质上讲是与电话调查同义的，尽管有很多电话调查仍然使用纸和笔。迄今为止，除了在减少数据丢失方面外，没有任何文献认为电话调查得到的数据，其质量不会受是否使用电脑辅助的影响（Catlin & Ingram，1988）。有关这方面的研究文献指出它的优点和缺点都与实际应用有关。它的主要优点包括便于问题形式和顺序的编排、数据输入速度快、样本管理方便，以及可用指令或所需定义的形式为访谈员提供“帮助”等。缺点是它需要一个准确无误的程序、访谈员难以对前面发现的错误进行修正及存在电脑系统崩溃的风险等。另外，虽然叙述式问题可以通过由访谈员逐字打出答案的方式予以处理，但是电脑辅助数据收集法，在更多的时候还是要求使用有固定答案项的问题。

大多数电脑辅助访谈都是通过一个集中电话设备进行的。

然而轻便的手提电脑的出现使访谈员用电脑辅助数据收集法进行个人家庭访谈成为可能。另外,在诸如医生办公室等这样精心选择的环境里,人们可以通过电脑自填问卷的方式收集资料:回答者坐在电脑前,阅读屏幕上的问题,然后在没有访谈员的情况下自行输入答案。那种有触摸式屏幕或者有鼠标功能的电脑特别适用于这样的资料收集方式。电脑辅助个人访谈(CAPI)的优点与电脑辅助电话访谈(CATI)几乎一样:便于问题的处理和资料的快速编辑。一个访谈日结束后,访谈员可以通过电话线传送资料以便集中保存处理。

尽管电脑辅助自填问卷调查的技术还在发展中,但其中还有更多令人感兴趣的潜力将被人们认识到。例如,通过电脑我们可以使用有别于文字形式的形式来呈现信息和刺激物(例如,图画)。电脑不但可以为那些有一定阅读障碍的人读出问题,还可以调整问题的语言以适应回答者的语言。在具有复杂的逻辑跳跃指令的自填问卷调查中电脑具有独特的优势,因为它可以改变问题的选择或次序,以适应先前给出的答案。交互式电脑可以提问,并可用按键法记录答案,这就为网上调查提供了一种新的选择方式,使回答者能在他们觉得方便的任何时候提供资料。最后,回答者在回答敏感性问题时会觉得输入电脑比面对面访谈更自在。

显然,电脑的发展,使得它成本更低、携带更方便和功能更强大。这一切将为研究者提供一种收集资料的新资源。尽管最为简单的使用方式在于电话访谈(这是当前最先进的),但是,电脑运用于家庭访谈和自填问卷调查(无论亲自或是通过网络)将是在今后十年内发展最迅速的领域。就像任何调查设计的特点一样,电脑辅助固然能解决许多问题,但它也会产生一些新问题。究竟哪种设计才是最好的,最终还是要取决于特定的调查需要。

各种方法的简要比较

前面的讨论并非详尽无遗，但是它涵盖了大多数的主要的调查方式。数据收集方式的选择是非常复杂的，它牵涉到调查研究过程的很多方面。接下来将总结各种数据收集方法的一些优缺点。

个人访谈的优点

- 有些样本只有通过个人访谈才能得到最好的效果（例如，区域概率样本）。
- 对于大多数总体来说，个人访谈可能是取得合作支持的最有效的途径。
- 访谈员具有处理问题的优势，诸如回答回答者的问题、探寻充分的回答、准确执行复杂的指令和顺序等。
- 可以使用多种方式结合进行数据收集，包括使用笔录和电脑的观察、视觉提示和自我填答。
- 可以建立和谐和信任的气氛（包括为获得某些非常敏感的资料而写的书面保证）。
- 可以使用比其他方式更长的工具（问卷）。

个人访谈的缺点

- 与其他方式相比，成本相对较高。
- 需要在样本附近地区有训练有素的访谈人员。
- 资料收集的总时间可能比电话访谈的方式长。
- 有些样本可能不如通过其他方式更易接近（如那些住在超高层建筑或高犯罪率地区的人、社会名流、职员、学生等）。

电话访谈的优点

- 与个人访谈相比，成本更低。

- 可以对一般总体进行随机数码拨号抽样。
- 与个人访谈相比,更易于接近特定的群体。
- 数据收集的时间较短。
- 有利于访谈员对访谈过程进行控制(与邮件调查或互联网调查相比)。
- 与个人访谈相比,对访谈员的雇佣和管理更容易:所需工作人员更少,不一定要在样本附近,并且能更好地进行指导和质量控制。
- 与邮件数据收集方式相比,回答率更高。

电话调查的缺点

- 抽样具有局限性,特别是排除了那些没有电话或者因为没有获得正确的电话号码而无法联系到的人。
- 与个人访谈相比,RDD 抽样产生的无回答率要高。
- 问卷或测量有很多限制,包括答案的可选择性、可视辅助的使用和访谈员的观察等。
- 可能不大适合涉及隐私或敏感性问题的调查。

自填问卷法的优点(与有访谈员参与的访谈相比较)

- 易于利用可视辅助来呈现问题(与电话访谈相比)。
- 可以使用有较长或较复杂的答案选项类别的问题。
- 可以问一系列类似的问题。
- 回答者可以不受访谈员影响而独立回答问题。

自填问卷法的缺点

- 需要特别仔细的问卷设计。
- 对开放型问题不是很有效。
- 回答者需要有较好的阅读和书写能力。
- 访谈员不能在实地对质量进行控制,如是否回答了全部问题、答案是否符合规定的要求、答案的质量等各个方面进行控制。

- 不能对由谁来回答问题进行控制。

自填问卷可以用于邮寄问卷、团体调查或者在家庭中进行调查等各种方式的调查。每一种方式都有其优点和缺点。

团体调查法的优缺点

- 通常具有较高的合作率。
- 有机会解释所研究的项目和回答与问卷有关的问题（与邮件调查相比）。
- 成本很低。
- 主要的缺点是：只有少数的调查能够使用团体样本。

邮件调查法的优点

- 相对来说成本较低。
- 可以在极有限的工作人员和设备的情况下完成。
- 可以取得分散的和很难通过电话调查或面谈方式联系到的样本。
- 回答者有时间仔细思考答案、查阅记录、询问他人。

邮件调查法的缺点

- 把邮件作为一种获取合作的方法几乎是无效的（取决于所研究的组群和主题）。
- 存在那些在数据收集过程中没有访谈员参与的各种缺点。
- 需要有样本的详细地址。

将问卷散发到户（以后再回收）式调查的优点

- 与邮件调查法相比，访谈员可以对研究项目作出解释、回答提出的问题，并指定户中的回答者。
- 与个人访谈调查的回答率几乎相同。
- 与个人访谈和电话访谈相比，回答者能更仔细地思考答案，并且有更多的机会查询记录或询问家人。

- 不需要训练有素的访谈人员。

将问卷散发到户（以后再回收）式调查的缺点

- 这种方法的成本与个人访谈的相当。
- 需要实地工作人员（尽管不要求像个人访谈那样经过精心训练）。

互联网调查的优点

- 收集数据的单位成本低。
- 回收速度快。
- 具有自填问卷法的所有优点。
- 具有电脑辅助法的所有优点。
- 像邮件调查法一样，为回答者思考答案、查找记录或询问他人提供了足够的时间。

互联网调查的缺点

- 样本局限于互联网的使用者。
- 需要详细的地址。
- 面临获得合作的挑战（取决于被抽样到的群体和研究主题）。
- 具有没有访谈员参与的数据收集方法的各种缺点。

最后，在考虑选择什么样的数据收集方式时，研究者也可以考虑把各种方式结合起来。如前所述，很多问题的答案不会受数据收集方式的影响。把个人、电话、邮寄和互联网资料收集方法结合在一起可能会节约总成本，因为那些成本比较低的模式不再需要原来可能需要的抽样或解决无回答问题的费用。迪尔曼（Dillman，2000）对有关结合使用的问题做了一些讨论。

结 论

显然,数据收集方法的选择是一个复杂的决策,它在很大程度上取决于特定的研究情况。所有提到的策略对某些研究而言是最好的,然而需要注意的是,在一般以户为基础的样本的调查中,情况会有所不同。25 年前,有研究者可能会认为个人访谈调查是大多数研究选择的方法,但是,这种观点难以说服那种认为别的方法也能获得令人满意的资料的争辩者。

研究者必须直接说明不能用电话访谈取代访谈员访谈的原因在于成本优势。尤其出于对主题、测量内容和无回答问题的考虑,研究者认为有些研究更宜于采用个人访谈法。然而,公正地说,这是研究者在解释不用电话调查的理由,而没有解释采用电话调查的理由。

自我填答技术在近十年里获得发展的原因有二:首先,互联网的发展为数据收集开辟了一种新方法。研究者期待对互联网潜能的进一步开发。其次,相当多的研究证明,自填问卷方法,尤其是有电脑辅助的方法,与有访谈员参与的方法相比,可以更好地收集到有关敏感话题的资料。这些发现伴随着研究范围的拓展,如毒品使用和危险的性行为,使人们产生了整合这些数据收集方法的兴趣。

最后,应当明确,在确定数据收集方法时,整体的调查设计方法是很重要的。与花费了同样的代价的较大规模的电话访谈样本相比,较小规模的个人访谈样本有可能提供更有用的数据集。在决定如何收集调查数据之前,树立一个正确的方法目标观,并对影响成本和资料质量的所有设计方面的问题进行缜密考虑是非常必要的。

练　习

不考虑金钱成本,设计一个使用邮件方式为最佳选择的调查,并说明,为什么使用邮件法比其他方法好。然后以同样的方式设计随机数码调查、个人访谈住户调查和互联网调查。

第5章　把问题设计成一个好的测量工具

在调查中,答案令人感兴趣不仅仅是因为答案本身,而且也因为它们与被测量对象有联系。好的问题有较高的信度(在类似的环境中提供一致的测量数据)和效度(答案符合问题所要测量的目的)。本章将从理论和实践两个方面探讨设计有信度和效度的问题的方法。

为某个调查工具设计问题实际是在设计一种测量工具，而不是查问用的问题。一般说来，对某一调查的问题给出的答案本身并没有什么令人感兴趣之处，答案之所以有一定的价值是因为它能在一定程度上显示出，它与一个令人感兴趣的客观事实或主观状态之间存在着的某种可预测的关系。好的问题能最大限度地增加记录的答案与研究者试图测量的内容之间关系。

从某种意义上说，答案只是由研究者精心设计的人为环境所引起的一些反应。本章要讨论的关键问题是，对于我们所关注的现实，一个调查问题的答案能告诉我们一些什么。现在我们来看几种答案和它们含义：

- 一个回答者告诉我们，在 1996 年的总统选举中他投了多尔（Dole）的票而不是克林顿（Clinton）的。我们实际感兴趣的是在投票站他拉动了哪一个控制杆。由于种种原因，调查的答案可能会与投票站中实际发生的事有很大的差异。回答者可能会拉动错误的控制杆，因而不知道他选举的到底是谁；回答者也有可能忘记了他投了谁的票；回答者也许由于某些原因而故意改变自己的答案。
- 一个回答者告诉我们在过去一年里他去医院接受治疗的次数。这与研究者在过去一年里每天 24 小时跟踪得来的数字是一样的吗？回忆本身的问题、看医生的定义问题、作出准确报告的意愿问题，都会影响到回答者所给出的次数与研究者独立得到的次数之间的一致性。
- 当一位回答者给公立学校制度的评价是“好”而不是“一般”或者“差”时，研究者希望能把这个答案用作学校制度的评价和认识的反映。假如回答者评论的是一所学校（而不是整个学校制度）、只是为了讨好访谈员而选择，或者对该问题的理解与他人不同，那么，他的答案就可能没有反映研究者试图测量的感受。

大量调查研究被人们分析和阐述似乎说明研究者的确能

了解答案的确切含义。旨在评估回答者的答案和真值之间一致性的研究表明，许多回答者能很好地回答大多数问题。尽管如此，那种假设人们给出的答案和事实完全一致的想法却未免过于天真。答案能成为一种很好的量度，通常都是精心设计的结果。下面我们要讨论的是各种研究者用来提高回答者的答案和事物的真实状态之间的一致性的具体的方法。

良好测量的目的之一是提高问题的信度。在两个回答人处在相同的环境时，他们应该会用相同的方式来回答问题。而实际上，由于随机误差的引进和测量并非完全精确，二者之间总会在一定程度上存在某种不一致。本章的第一部分将讨论怎样增加问题的信度，还要讨论给出的答案与研究者所试图测量的目的之间的关系问题：答案的一致程度有多高？本章的最后两部分会谈到效度问题，答案和真值之间的一致性问题，和提高一致性的方法问题（Cronbach & Meehl，1955）。

提高答案的信度

确保测量一致性的一个措施是使样本中的每个人回答相同的问题，并将答案记录下来。研究者会期望做出这样的假设：答案差异来自回答者的差异，而非来自每个回答面前的刺激物的差异。显然，问题的措辞是刺激物的主要部分。

调查数据的收集是研究者和回答者之间的一个互动过程。在自填问卷的调查中，研究者通过书面问卷或电脑屏幕与回答者进行直接交流；在其他方式的调查中，访谈员为回答者读出问题。在每个调查中，调查手段都是为调查对象提供问题。为了使所有的回答者具有相同的数据收集经历，好的问题应具有以下特点：

- 对问答过程中的研究者一方而言，他们要做的工作是照本宣科，所以他们必须事先准备好书面问题，以备回答人对

它做出回答。

- 对于每个回答者来说,问题的含义都是相同的。
- 在向所有回答人传达构成恰当回答问题的各种答案时,应该完全一致。

不恰当的措辞

问题的措辞不当的最简单的例子是研究者使用的措辞不能构成一个完整的问题。

措辞不完整的例子

	不好的	较好
例 5.1	年龄?	在过上一个生日时,你的年龄多大?

访谈员(或回答者)必须自己加上一些词或者重新措辞,以使上面左边的问题便于回答。如果目标是让所有的回答者都回答同样的问题,那么研究者最好是将问题完整写出来。

有时需要在问题中使用可选择式的措辞,以使问题适合不同回答者的不同的具体情况,但这并不意味着研究者不必写出问题。一般是把可选择的措辞放在括号里,以备访谈员在需要时使用,而在不需要时则可略过。

选择性措辞的例子

例 5.2　在过去一年里,你(或任何一个和你住在一起的人)曾经被陌生人攻击或殴打过吗?

例 5.3　(你/他/她)向警察报过案了吗?

例 5.4　在(你/他/她)的上一个生日(调查户中的每一个人)的年纪是多大?

在例 5.2 中,如果研究者已经知道回答者是独居的,那么括号里的插入语就会被省略。相反,研究者将保留它。在例

5.3里，括号内的选择看起来是不重要的，但是，它却提醒访谈员必须对用词作出选择使用正确的代词。而这里必须坚持的原则是，访谈员一定要完全按照文字，照本宣科地宣读问题，以能呈现一个令人满意的刺激。

例5.4是一个用一个某些措辞可以有所变化的问题完成同样的任务的例子。如果要用同样的问题对每一个家庭成员提问，就需要使用这样一种格式的问题。与其不断地重复同样的话，不如只写一个问题，并且指导访谈员用合适的词进行替换（你的丈夫/你的儿子/你的长女）。

当然，电脑辅助手段的一个优势是可以用电脑填入合适的词语，而不是由访谈员根据各种情况调整问题的措辞。不管是写在纸上还是通过电脑，目标都是让访谈员提出清楚的问题，并利用访谈员在以前的访谈中所获得的知识，根据回答者的不同情况对问题作出修改。有一种偶尔出现在问卷中的，但是我们不可接受的选择性措辞。

不可接受的选择性措辞的例子

例5.5　在你住所附近你最喜欢什么？（我们对任何事情都感兴趣，比如房子、人、公园或其他不论什么东西）

我们猜想，在问卷的编制者看来括号中附加的说明可能会对那些回答问题有困难的回答者有所帮助。但是从测量角度来看，它破坏了标准化访谈的原则。如果回答者不能立刻想出问题的答案时，访谈员就使用了附加说明，那么属于那个未能立即想好答案的回答者的子组（subset）的那些回答人，就可能回答的不是同一个问题。诸如这样的附加说明通常是在研究者认为最初的问题表述得不是很好时引进的。恰当的做法是一开始就写出含义明确的问题。在问题的哪一部分应该宣读，和怎么样宣读这些问题上不给访谈员任何选择的余地。告诉访谈员应该像刚才讨论过的例子那样，必须以标准化的方式使问题适合于一个特定的回答者的具体情况。

下面是又一个有关问题措辞不完整的例子，这个例子中有三个错误的地方：

措辞不当的例子

例 5.6　我希望你能通过使用非常好、好、一般或差这四个等级来评估你周围事物的不同特点。在我念出以下选项时请认真思考。

A　公立学校

B　公园

C　公共交通

D　其他

例 5.6 中存在的第一个问题是问题主干的顺序问题：选择项的阅读先于关于仔细思考的特定项目的引导语。回答者可能会忘记问题，因此访谈员可能不得不在回答者做出回答之前做一些解释或重复宣读。存在的第二个问题是，调查员对第二个选项中的"parks"一词没有加以限定（因为，在英语中，parks 有多种译法，如公园、停车场、运动场等，容易产生歧义——译者注）。更好的措辞应该像下面这样：

措辞得当的例子

例 5.6a　我想问你对周围事物的不同特点的评价。我希望你能仔细思考你的答案。你是怎么评价（需要评价的特点）的——很好、好、一般或者差？

这样的编排使访谈者首先问最需要问的问题，然后再列举所有的选项。

例子 5.6 中存在的第三个问题是，对于第 4 选项"其他"，访谈员会说什么？可以临时编造一些新的问题吗？例如"你还有对周围事物的其他评价吗？"这个不确定的问题如何表述？

类似例子中这种"其他"选项在问题单中出现的情况还比较普遍。显然,例 5.6 这种形式的问题表述是不恰当的。

上述例子说明:如果措辞不完整,就不能保证对所有回答者所做的提问是一致的。增加一致性的另一个步骤是编制一系列流畅而又简明的问题。如果问题使用的措辞令人费解或含混不清,有些词语很难读或者连成句时感觉很累赘,访谈员就要换一些词以使问题听上去更顺耳或者更方便读出。虽然可能通过对访谈员进行训练和指导也能达到这种要求,然而,最明智的办法还是尽可能给访谈员容易读的问题以便于他们提问。

确保所有的回答者对问题含义的理解一致

如果对所有的回答者问完全一样的问题,采取了这样的步骤可确保答案中的差异将归于回答者之间的差异。但除此之外,还有需要做进一步考虑的问题:对所有的回答者而言,问题的含义都应该是一样的。如果两个回答者对问题的理解不同,那么,他们的答案就可能因此而不同。

这样的问题之所以产生有可能是因为使用了不能为人们所普遍理解的词汇(对于一般的人口总体而言,重要的问题是我们必须牢记,各种学历和文化背景的人在样本中的代表的比例与总体是一致的)。即使面对的是受过良好教育的回答者,也应该使用那些简短和普遍都能理解的词语,了解这一点对问卷设计是非常重要的。

毫无疑问,比使用不熟悉的词语更容易犯的错误就是使用专业术语或者容易产生歧义的概念。对日常语汇的误解也是一个比较普通的问题,有些学者对这一问题作了详细的论述(例如,Belson;1981; Fowler,1992;Oksenberg, Cannell,& Kalton,1991;Tanur,1991)。

术语定义不明

例 5.7　去年,你看医生或者与医生讨论你的健康状况的次数

是多少?

这个例子中存在的问题是,有两个模糊不清的术语或概念:第一是医生的概念不明确。是不是只包括那些从事医药实践并具有医学博士学位的人?假如这样的话,精神病学家应该被包括在内,而心理学者、脊椎指压治疗者、整骨疗法家和足病医生则被排除在外。医生的助手或者在医生办公室直接为医生工作的护士算不算?如果一个人去医生的诊室接种疫苗,但是是由护士执行的,这种算不算在内?第二,去见医生或者与医生讲话的内容是什么?打电话咨询算不算在内?去见医生但是医生不在,这算不算?

解决这个例子中存在问题的最好办法,通常是向回答者和访谈员提供他们所欠缺的定义:

例 5.7a　我们将要询问的是你见医生和从医生那里得到的医疗建议。在这种情况下,我们感兴趣的医生是所有具有医学博士学位的人或者在办公室直接为他们工作的专业人员,例如护士或医疗助手。

正如在这个问题中所遇到的情况一样,当定义极其复杂并费时较多时,可采用另一种方法:对很可能被遗漏的事情提供补充性的提问。例如,对见精神病学家、接种疫苗以及电话咨询经常报告不够,这就需要确保有专门的后续问题作为补充。

术语定义不明

例 5.8　昨天你吃早餐了吗?

该例的问题在于早餐的定义相当宽泛:有的人认为在中午以前的任何时候喝的咖啡和吃的油炸面包圈都算是早餐;另外一些人则认为那不算吃过早餐,除非包括火腿和鸡蛋,并且要

在早上八点以前吃完。如果调查的目的是测量早晨的食品的消耗量，那么结果就可能会由于对早餐的定义不同而出现很大的误差。

解决方法有二：其一，是选择对早餐的概念进行定义。

例 5.8a　基于研究的目的，我们把早餐当作一顿正餐（meal），在早上 10 点以前吃；包括一些蛋白质的食品，例如鸡蛋、肉类或者牛奶；包括一些谷类食品，例如烤面包或者谷类早餐食品；以及一些水果或者蔬菜。根据这些定义，请问你昨天吃早餐了吗？

尽管给早餐下定义这的确是一种不错的方法，但是却太复杂。除此之外我们还可以有第二种方法。这种方法与第一种方法的区别在于，研究者不告诉回答人早餐的普遍定义，而只是要求他们报告上午 10 点以前他们消耗的具体食物。在编码阶段，才对这些食物用统一的标准进行评价，再确定他们吃的食品是否符合早餐的标准。这样做不需要回答人对早餐有相同的定义。

术语定义不当

例 5.9　你赞成还是反对枪支管制立法？

例 5.9 中存在的问题：枪支管制立法可能指在法律上禁止某些种类的枪支销售，要求人们对所持枪支者注册，限制人们可以拥有的枪支数量或种类，或者对什么人可以拥有枪支加以限制。如果不假定回答者对这个问题的理解，答案就无法解释。毫无疑问，回答者将会以不同方式阐释这个问题。

例 5.9a　有一个控制枪支的议案是这样的，不允许任何被证明有过暴力犯罪记录的人购买或拥有手枪、步枪或者鸟枪。你是反对还是赞同这样的立法呢？

有人可能会争辩说这只是枪支管制多种议案的一种。确实是这样。如果有人要问有关枪支管制的各种方法的复合问题,那么,他就应当把问题分解成多个具体的问题,以便能被所有回答者普遍理解以及能被研究者普遍解释。研究者不解决好这样的复杂问题而把它留给回答者自己来决定他们要回答的是什么问题,这样做是无济于事的。

在试图给所有回答者交代一个复杂的概念和力图能保持问题的简单明了二者之间存在一种潜在的张力。这种情况在访谈员参与的调查中特别突出,因为以口述的方式传达时,对一个很长的定义是难以把握的。

有一种可能的方法,是告知访谈员在回答者要求明确问题或对问题产生误解时才提供定义。值得注意的是,这种方法中如果定义是访谈员临时提供的,那么他们提供的定义能否前后一致。不过,电脑辅助访谈使得为访谈员提供精确的定义变得容易起来。用这种方法时,更需注意的另一个问题是,只有一部分回答者才得到了所需的定义,而那些没有要求明确问题或对问题没有显示出误解的回答者,会缺少那些很可能影响他们答案的重要信息。

在康拉德和斯科伯(Conrad and Schober,2000)的实验中,给访谈员根据回答者的需要提供定义和解释问题的自由。虽然有事实表明精确度确实提高了,可是随之而来的是访谈员培训费用的增加和访谈时间的加长。针对更多的研究需要解决的如何就复杂概念进行提问的状况,适用于大多数调查的最好方法是:避免使用复杂、模糊的术语,并对那些在问题表述中用到的术语进行定义。

避免出现复合问题

另一种导致问题不具有信度的情况是一题两问。

例 5.10　你希望自己富有并出名吗?

问题很明显:“富有”和“出名”是不同的,可能有的人只希望拥有其中的一个方面。当面对这样的两个问题时,回答者不得不自己决定回答哪一个,而且取舍也因回答者的不同而不同。

大多数的复合问题有些难以捉摸,如:

例 5.11　在过去的 30 天里,当你从 ATM 柜员机上取款时,你取款少于 25 美元的频率——总是、通常、有时、从不?

这个问题要求进行 3 次认知计算(cognitive calculation):计算去 ATM 柜员机上取款的次数,计算取款少于 25 美元的次数,以及两者之间的关系。尽管从严格的意义上来讲这只是一个问题,但是它需要预先回答两个问题才能得出答案。如果将其设计成两个问题将会好点。

例 5.11a　在过去的 30 天里,你从 ATM 柜员机上取款了多少次?

例 5.11b　(如果取过)取款少于 25 美元的有多少次?

例 5.11a 和例 5.11b 问题组的另两个优点值得注意:第一,它区分了那些从不使用 ATM 柜员机的人,因为这个问题不适合那些人。第二,两个问题都使用具体数字,避免了让回答者做计算的麻烦。简化对回答者的要求,几乎总是一种好的提问方法。

例 5.12　你到什么地方进行日常医疗保健?

这个问题假设所有的回答者都进行了日常医疗保健,但这

并非是一个精确的假设。它应该分成两个问题来问。最好的方式可能是:先问在某一时段是否进行了日常医疗保健(例如,过去的12个月里),如果回答是肯定的,再问在什么地方进行保健的。

"不知道(Don't Known)"选项

当回答者被问到有关他们自己的生活、感受或经历时,答案"不知道"经常表示他们不愿意回答要求他们回答的问题。另一方面,有时它也表示所提的问题他们确实不知道。在问题的主题远离他们当前的生活时,更加可能发生的情况是,有些被调查人常常会因为对问题的主题缺乏足够的知识,而无法形成一个见解或想法。这种情况实际上就是一题两问的又一种实例:你有回答这个问题所需的信息吗? 如果有,那么你的答案是什么?

有两种方法可以处理诸如这样的可能情况:一种是简单地直接向所有回答者提问,是否给出"不知道"这一答案取决于回答者的自愿。然而,回答者自愿回答"不知道"的意愿不尽相同(Schuman & Presser,1981),且访谈员处理"不知道"选项时的方法也不一致(Fowler & Mangione,1990;Groves,1989)。备择的方法是,问所有的回答者一个标准化的筛选性问题,即问他们是否对某个主题足够熟悉,以至能对它有一定的见解或想法。

当回答者对主题很熟悉时,是否要进行筛选性提问并不重要。当有相当数量的回答者不熟悉或没有想过那个主题时,不管提问的是什么问题,最好要先对相关主题做一个熟悉程度的筛选性提问。

对特定的子群使用特定的措辞

研究者一直在致力于实现对不同的子群使用不同的词汇。我们可以证明标准化测量需要对不同的子群使用不同的问题(Schaeffer,1992)。

而为不同的子群设计不同形式的问卷几乎是不可能做到的。因而方法论学家都在设法找到对于整个总体含义都一致的措辞。即使在遇有一个问题的措辞对一部分人（大多是对具有良好的教育那部分人）更具有语言上的典型性这样的情况时，要找到与之完全相当的适合另一部分人措辞，并作出明确的规定，使访谈员在提问时可据之确定在什么时候用哪一种版本的问题是十分困难的。不仅如此，这样做的结果常常还会是得不偿失的，因为两套版本的问题带来的不可靠性常常大于它所能消除的不可靠性。

最大的挑战是怎样从说不同语言的人群里收集具有可比性的数据。最为认真的做法是，找一位翻译把原文翻译成另一种语言的新版本，然后让另一位翻译把新版本回译成原来的语言，再尝试消除原文和回译后得到的版本之间的不同。

如果问题的设计者最初就考虑到要如何便于翻译的问题，那么这个过程就能大大加快。例如，对数字的翻译就比对形容词的翻译容易。想要把抽象的概念和口头语翻译准确就十分困难。即使十分认真地进行翻译，也很难保证人们用不同语言回答的问题具有可比性。不同的语言中，形容词形容的程度的可比性是令人怀疑的。问题越具体，不同的语言和文化间的可比性就越高。马林 G. 和马林 B. V.（Marin & Marin，1991）很好地分析了从讲英语和讲西班牙语的人那里收集可比资料所面临的挑战。

规范答案类型

如上所述，给访谈员一个好的文稿以便他们能照本宣科地读出问题，而这一点对使设计的问题让所有的回答者的理解都一致是很重要的。好的问题还有一个关键组成部分，那就是它能使回答者对由什么构成问题的恰当答案这一点达成共识。

让回答者在什么构成了问题的恰当答案这一点上达成共识的最简单的办法是给他们提供一张可以接受的答案的清单。这样的问题叫作封闭型问题。回答者必须从研究者提供的一

组备择答案中选出一个,或多个答案。

封闭型问题并不适合所有的情况。与提供的答案相比,问题本身的答案可能范围更加广泛,但这些答案并不是都可以预测。因此,研究者会倾向于不给回答者提供可选择的答案。这时,提出的问题就必须尽可能地交代清楚希望得到的答案是什么。

例 5.13　你什么时候得的麻疹?

例 5.13 中存在的不足是,这个问题没有对回答者需要作出回答的时间一词作出明确规定。因此下面这样的答案都是可能的:“5 年前”“当我在部队的时候”“当我怀上第一个孩子的时候”“当我 32 岁的时候”“在 1987 年”。所有这些答案都可能是同一个人给出的,而且对上面这样一种形式的问题这些答案都是可以被看作是恰当的。但是在同一个调查里,这些答案并非都是可以接受的,因为描述性统计要求具有可比性的答案。访谈员使用例 5.13 里边的措辞,不能始终都得到可比数据,因为每个回答者都必须猜测调查想要的答案究竟是什么。

解决上述问题的办法是,重新编写问题,给回答者解释清楚需要的答案是什么样的。

例 5.13a　你在多大的时候得过麻疹?

很明显,像例 5.13a 使用的措辞方式,才是一个研究者准备向所有的回答者提出的问题时应该使用的措辞方式。

例 5.14　你为什么投候选人 A 的票?

几乎所有问“为什么”的问题都会引起这样或那样的问题。因为人们对因果问题的认识和参考的框架都可能对答案

产生影响。在这里讨论的这一特定的例子里，回答者可能会选择谈论候选人 A 的优势、候选人 B 的弱势，或者谈其选择某种标准的理由（如，我的妈妈一生都是共和党人）。因此，即使是那些观点一致的回答者，他们给出的答案也可能不同。

解决这个问题的办法是明确问题关注的焦点：

例 5.14a　候选人 A 的哪些特点让你投他/她的票，而不是投候选人 B 的票？

一个采用了诸如这样的措辞的问题，本身就向回答者阐明了研究者希望他们谈一下，他们投票选举的候选人 A 的情况。如果所有回答者的回答都是以同样的参考框架进行的，那么研究者就可以对来自不同回答者的答案，以直接的方式进行比较。

例 5.15　你最喜欢你居住地周围的哪些东西？

在回答这样的问题时，有些人可能只列举一或两点，而另一些人则可能会列举很多。诸如这样的差异有可能反映了回答者在感知或情感上存在的重大差异。而研究已经清楚的证明，教育程度与对诸如这样的问题给出的答案数是高度相关的。另外，访谈员对答案的数目也有影响。

解决的方法是，在问题中明确需要给出的答案的数目：

例 5.15a　请告诉我们，在你居住地周围你最喜欢的一个事物是什么，它的特点是什么？

例 5.15b　请告诉我们，在你所居住地周围，你最喜欢的三样东西。

虽然这样的方法并不能完全令人满意地解决所有类似的

问题,但对于许多类似这样的问题而言,这种方法的确是减少不同回答者之间的答案中存在的不必要的差异的一个有效途径。

这种方法最基本的一点是:因为回答者可以根据要求选择适合自己的答案,所以答案可以是因人而异的。正因为这样,尽可能清楚地阐明我们所需要的答案的性质,便有可能去除量度过程中的不必要的不可靠性的来源。

测量的类型/问题的类型

引　言

这些程序被设计用来最大限度地提高研究的信度,即尽可能地使身处可比情景中的人们能以相似的方式回答问题。尽管人们可以进行高信度的测量,然而测量的结果仍然有可能并非总如我们所愿。给出的答案作为一个真的量度的程度,以及它体现研究者希望或期待它体现的东西的体现程度叫做效度。这一节将就问题设计的某些方面展开讨论,讨论的内容除了如何进一步提高问题的信度外,还包括如何提高调查测量的效度。

在进行这样的讨论时,我们有必要区分两类不同的问题:一类是用来测量事实或客观事件的问题;另一类是用来测量主观状态诸如态度、意见或情感的问题。尽管有些问题介于这两类问题之间,但是效度概念在客观性测量和主观性测量上还是有所不同的。

如果有可能通过一些独立的观察来检查答案的准确性,那么效度的测量就变成了有关与某种"真(true)"测量值的相似性的调查报告。从理论上说,人们可以得到独立的、准确的关于一年内某人使用 ATM 柜员机的次数。尽管实际上要得到这样的测量十分困难(例如:不可能获得相关记录),但是,我们可

以把效度理解为与客观情况的一致。

相反，当人们被问到一些主观状态、情感、态度和意见时，我们无法采用客观的方法对答案加以确认。只有回答者本人才清楚自己的感情和观点。正因为如此，主观状态的效度报告只能通过间接的办法作出，如对某人给出的与之关联的其他答案，或对人们认为与量度的主观状态关联的有关回答人生活中的某些事实进行评估作出对于诸如这样的测量而言，做到真正独立直接测量是不可能的，答案的意义必须从关联的模式中推断出来。

测量的层次

在社会科学中有 4 种不同的测量方法。这就产生了四种不同的需要回答人完成的任务，和四种不同的需要进行分析的数据：

- **定类**(nominal)：人或事件被分成了无序的类别(如：你是男性还是女性)。
- **定序**(ordinal)：人或事件是有序，或按照单一的维度被分成了有序的类别(如：你怎样评价你的健康状况——非常好、好、一般还是不好)。
- **定距数据**(interval data)：附加于类别的数字提供了有关有序刺激物或等级之间的距离的有意义的信息(事实上定距数据是非常罕见的，华氏温度是为数不多的常见的例子之一)。
- **定比数据**(ratio data)：指定的数字不仅可以确定数字间的间距，而且可以计算这些值之间的比率。常见的例子是用某种客观的物理尺度如距离、重量或者压力测得的计数或测量值(过了上一个生日后你多少岁了?)。

收集事实调查资料时，经常要求回答者把他们自己或自己的经历归入某一类，建立一种定类数据，或问他们一个数字，这样建立的数据大多是定比数据。“你有工作吗?”“你结婚了

吗?”“你有关节炎吗?”这些问题都是提供定类数据的例子;“你看过多少次病?”“你的体重是多少?”“你每小时的工资是多少?”,这些例子中的问题,都是要求回答者为定比数据提供真实数字。

在收集事实时,也许也会要求回答者提供定序的答案。例如,回答者被要求以比较大的类别来报告他们的收入,或者要求他们用非数值的言辞描述他们的行为(例如:经常、偶尔、极少或从不)。回答者之所以被要求用定序语汇报告实际事件,通常是因为研究者并不需要答案十分精确,或认为报告确切的数字可能过于困难。然而,一个事实问题的定序答案的构成,通常都是以真正的数字为依据的。

至于主观性数据的报告情况则有所不同。尽管经过了多年的努力,心理物理学心理学家(psycho-physical psychologist)(例如:Thurstone & Chave,1929)首先对人们的主观状态赋以数值,以满足定距和定比数据的假设,但是,在大多数情况下,还是要求回答者提供关于主观状态的定类和定序数据。定类问题是“你的情感、意见或者感觉是属于哪一类的?”;定序问题是“你的情感、意见或者感觉处于这个连续统的哪个位置?”。

在设计调查工具时,研究者的一项基本工作是决定所希望的使用测量的种类。这样的决定做出之后,问题的形式便不言自明。

问题的类型

调查问题可以大致分为两类:那些为回答者提供可以接受的答案的清单的问题(封闭式问题)和那些没有为回答者提供可以接受的答案的清单的问题(开放式问题)。

当目的是要把人们归于无序类别时(定类资料),研究者既可以选择提出封闭式问题,也可以选择提出开放式问题。实质上相同的问题可以用二者中的任何一种形式设计。

开放式和封闭式问题的例子

例 5.16　你的健康状况如何？（开放式）

例 5.16a　你平常的健康状况是以下哪种情况？（读选项，封闭式）

例 5.17　你认为我们国家目前遇到的最重要的问题是什么？（开放式）

例 5.17a　这里有一个很多公民都关心的问题的清单。你认为目前国家遇到的最大问题是其中的哪项？（封闭式）

开放式问题有一些优点：首先，开放式问题使研究者可以得到意想不到的资料；其次，开放式问题更接近回答者的真实想法；第三，这点很重要，回答者喜欢有机会用自己的语言回答问题，如果只是从所提供的选项中选择答案，没有用自己的语言回答问题的机会，他们会很失望；最后，当可能的答案清单长得难以向回答者提供时，用开放式问题则更为适合。

除了不具备上述优点，封闭式问题通常是一种更令人满意的建立数据的方式。原因有以下三点：

- 当给出选项时，回答者能更可靠地完成问题的回答。
- 当给出选项时，研究者能更好地解释答案的含义（Schuman & Presser，1981）。
- 当提出一个完全开放式的问题时，许多人给出的是一些比较少见的无法作统计分析的答案。给回答者提供数目有一定限制的答案选项，将会提高有足够的人给出的任何特定的答案用来进行统计分析的可能性。

最后，如果研究者希望得到定序数据，就要给回答者提供类别。除非一组允许的有序答案已经在问题中做了明确的规定，否则便无法沿一单独的连续统对答案进行可靠的排序。当要求回答者完成一个定序任务时，我们提倡与他们就要完成的任务做进一步讨论，因为定序量度可能是抽样调查中使用最为普遍的一种量度类型。

表 5.1 给出了一个连续统（虽然这个例子只是让回答者给

某种类别标上等级,但这种方法具有普遍性,适用于所有定序问题)的例子,其中有一个研究者假设的维度,这一维度涵盖从最消极的感情可能到最积极的感情可能。研究者将回答者归入有序类别的方法,是在这样的连续统上标上名称或标签。然后,要求回答者考虑这些标签、自己的感受或观点,并将自己归入适当的类别。

关于从诸如这样的问题中获得的各种数据,有两点值得指出:第一,回答者可能对标签或者类型的意义有自己不同的理解。为了进行有意义的分析,唯一需要假定的是:通常,那些报告自己感觉"好"的人比那些报告自己感觉"一般"的人更积极。尽管回答者对"好"和"一般"的理解,和对"好"和"一般"的标准总是多少会有所不同,因而使测量有不可信之处,但是测量仍然还是有意义的(即与研究者希望测量的潜在感觉状态相关)。

第二,像这样的一个定序量表的测量是相对的,人们选择特定标签或类型的分布状况,取决于研究者所提供的特定量表。

再来考虑表 5.1 中的等级量表和两种创建定序量表的方法。一种情况是,研究者使用的是三级量表:好、一般和不好。另一种情况是,研究者使用了五级量表:极其好、非常好、好、一般和不好。比较两种量表我们可以发现,后者增加的"极其好"、"非常好"决不会是简单地把"好"分成三部分,相反,它改变了量表的整个感觉。人们在对类别的相对位置作出回答的同时,也对与之对应的描述作出了回答。处于量表中第 4 点的"一般"比处于量表第 2 点的"一般"更接近连续统上的消极端。因此,我们可以预期,用 5 级量表将比用 3 级量表得到更多的"好"。

表 5.1　对某事的感觉

非常积极的				非常消极的
两类量表(two-category scale)				
好				不好
三类的量表(three-category scale)				
好		一般		不好
四类的量表(four-category scale)				
非常好	好	一般		不好
五类的量表(five-category scale)				
极其好	非常好	好	一般	不好

如果能将这些量表按它们假设的那样来对人们进行排序，那么这些量表是有意义的。然而，抛开假设，只是就量表本身作出总体的某个比例觉得某事“好或较好”的陈述是不恰当的，因为这意味着总体是在某种绝对意义上被描述的。事实上，如果问题不同，比例就会改变。在使用定序测量时，只有比较性陈述(或对关系的陈述)才是正当的：

- 在组群之间对同一问题的答案的比较(例如：A 组里认为候选人“好或者更好”的人数比 B 组多 20%)。
- 对可比样本的答案的跨时间比较(例如：在 1 月份认为候选人“好”的人数比 11 月份的人数多了 10%)。

在将相同的普遍性的评论应用于让回答者给选项排序而获得的数据(例如：在学校、警察局和垃圾处理站中，你认为哪一个是最重要的城市服务设施？)时，将一个选项排在首位或中间的百分数完全取决于所提供的特定选项清单，因此在那些备择项完全改变时，分布状况之间的比较就毫无意义了。

同意—不同意选项：一个特例

在调查研究中同意—不同意选项的使用十分普遍，因此应

该予以特别的关注。在诸如这样的一些选项里，回答者要完成的任务不同于他们在将自己归入一个定序的类别时遇到的任务。通常的做法是给回答者作一个陈述，然后问他们是否同意那个陈述。陈述被置于如表5.1所示的那样的连续统中的某处。通过弄清楚回答者是否说自己的感觉十分接近那一陈述（通过说同意），或感觉自己与那一陈述相去甚远（通过说不同意）来算出他们在连续统上的位置。

当有人将同意—不同意式的问题与直接定级式的问题进行比较时，就会发现前者有很多缺点。比较下面两种问题形式的例子：

例5.18　我的健康状况不好。你是完全同意、同意、不同意还是完全不同意呢？

例5.18a　你是怎样评价你的健康状况的——极其好、非常好、好、一般或是不好？

第一种陈述形式有以下几个缺点：

- 等级量表把回答者分为五类，而同意—不同意问题的分析几乎总是把回答者分为两个组（同意的人或不同意的人）。因此，用等级量表可以得到更多的信息。
- 为了方便解释，同意—不同意问题只能问及一个连续统的两个极端。如果一个人做出的陈述是“我的健康状况一般”，它既不是“好”也不是“不好”，他就无法在这样的连续统中将自己归类。这一性质限制了量表的排序能力，因为它无法将人们置于连续统的中间。
- 回答者常常会对以这样一种方式提出的问题感到迷惑不解，因为他们认为同意自己身体健康的说法，并不一定等于不同意自己身体不健康的说法。
- 研究表明，当问题以这样的方式提出时，有些回答者特别喜欢说“同意”（或勉强同意）。就是说，如果问题5.18以

否定的形式提出，那么有些人就会既同意他们的身体是“不健康”，也同意他们的身体“不是不健康”（Dillman & Tarnai，1991；Schuman & Presser，1981）。

对于一维测量任务来说，难以证明使用例 5.18 比使用例 5.18a 更合理。但是，在获得诸如下面这样的复杂陈述的回答时，这种形式则使用得非常普遍：

例 5.19　考虑到目前的经济条件，有两个以上的孩子是不合理的。

这个问题同时问了三种事情：对经济状况的认识、对合理的最多孩子数的认识和对经济与家庭规模的关系的了解。

例子中存在的问题：如果一个人没有想到经济条件糟糕（这一点在问题中只是一个假设），也没有认识到经济条件与家庭规模的任何关系，他碰巧却认为一个家庭有两个孩子最合适，那么，他回答这个问题就很难。此外，不管一个人是说同意或者说不同意，都很难知道他同意或者不同意的是什么。

同意—不同意形式似乎是设计问题的一种相当简单的方法，而事实上，要用这种方式进行可靠、有用的测量是困难的，并且需要非常小心谨慎。通常，如果研究者避免使用同意—不同意的问题方式，就能得到更可靠、更有效和能够解释的数据。

提高事实报告的效度

当研究者提出一个关于回答者的某个事实的问题时，其目的是让回答者给出十分精确的报告。也就是说，研究者希望回答者给出的答案应该与研究者个人（假如研究者有办法得到回答这一问题所需的信息时）给出的答案相同。目前有大量的关于事实材料报告的方法论方面的文献资料。在各个领域，尤其

在经济和健康领域,研究者一直在将调查得到的报告与各种事实记录进行比较(参看:Cannell, Marquis, & Laurent, 1977; Edwards et al., 1994; Edwards, Winn, & Collins, 1996)。

回答者准确地回答了很多问题。例如,由调查报告得悉,在调查涉及的六个月内,有90%的人报告说曾经住过院(Cannell et al., 1977)。而人们报告的质量既取决于对他们提问的内容也取决于对他们提问的方式。回答者报告的事件不太准确的基本原因有以下四个:

- 他们没有明白问题的含义。
- 他们不知道问题的答案。
- 他们理解问题,但是想不起来了。
- 他们不想在访谈情景中回答。

研究者可以通过一些措施克服这些潜在的问题。下面对这些措施进行讨论。

理解问题

如果回答者对提出的问题有不同的理解,就肯定会产生误差。如前所述,当研究者试图计数具有复杂定义的事件,如入室行窃次数或者看病次数时,他们有两种选择:①给所有回答者提供定义;②让回答者提供对他们的经历进行详细的、复杂的分类所需要的信息,再让编码员对答案进行分类编码。

福勒(Fowler, 1992)的研究表明,人们在回答含有含义不清的语汇的问题时,会产生颇为失真的数据。研究者不能假设回答者在不明白问题的含义时,就会去询问研究者。为了尽可能地提高事实调查数据的效度,最基本的、也是首要的一步是编写能被所有回答者理解的问题。

缺乏知识

作为误差来源之一的缺乏知识有两种主要类型:①抽到的回答者不知道问题的答案,但是在抽到的家庭里可能有其他家

庭成员知道；②在抽到的家庭里没有一个人知道答案。解决第一种情况的方法在于选择正确的回答者，而不在于问题设计。在很多时候，问题出在要求户内的一个回答者报告有关其他家庭成员，或整个家庭的信息上。解决办法包括以下几点：

- 确定并对最了解情况的家庭成员进行访谈。
- 使用的资料收集方法要给回答者有向其他家庭成员咨询的机会。
- 不使用代理回答者，回答者只提供有关自己的资料。

有时需要采用一种复杂的资料收集方法。例如，人口普查局进行的国家犯罪调查，虽然从单独一个家庭的情况报告者那里得到整个家庭的犯罪的报告，但此外，也要直接询问每个家庭成年人个人的犯罪，如抢劫的情况。如果访谈基于面对面的方式，那么便可在访谈完成之后，给那些不在家的成员留下自填的问卷，或通过电话再进行第二次访谈，这样与家庭其他成员的访谈成本就会减少。有一种变通方式是让主要的回答者尽可能地报告所需的有关所有家庭成员的资料，然后给回答者寄出一个让其核实的概要，同时允许他与其他家庭成员商量。

当回答者答不出与自身有关的问题时，那就是问题设计上存在缺陷。从理论上来说，研究者能够区分开是回答者回忆不起来的信息还是他们根本就没有的信息。在任何情况下，研究者要做的是设计出几乎每个人都能回答的问题。而为了达到这一目的，可供我们选择的方法有以下几种：

- 修改问题，使询问的信息不含太多的细节，或比较容易回忆。
- 帮助回答者对答案进行评估。
- 改变或者放弃这一目标。

经常有一些问题要求答案的详细程度超过了研究目标的需要：

当研究目标是为了找出哪些人因为高血压而吃药时

(一个比较简单的问题),提出的问题要求回答者说出他们使用的所有药物名称(一个很难回答的问题)。

当只需估计出收入的大概类别就能满足研究目的时,提出的问题却是开放式的(因而意味着要很详细)。

回忆会遵循几个很明显的原则:不很重要的小事比那些具有重要意义的事情更容易被遗忘;报告最近发生的事情比报告很久以前所发生的事情会好一些(Cannell, Marquis, & Laurent, 1977)。有时可以通过提问更容易回忆的问题来提高报告的质量,这种改变问题目标的做法是很有意义的。例如,尽管期望回答者能报告过去一年发生的所有犯罪事件,但是如果研究者只要求报告最近6个月的,报告中的误差就会更小。

随着越来越多的认知心理学家加入到了调查方法研究的行列,一系列新的问题设计策略应运而生(Jabine, Straf, Tanur, & Tourangeau, 1984; Sirken et al., 1999; Schwartz & Sudman, 1996)。各种策略都致力于帮助回答者回忆所发生的事件(例如:通过提出可能的联想)或者把事情放到一定的时间背景中(例如:让回答者回忆发生在一年前的某些事)。研究表明,对于很多调查任务来说,回答者并不真正通过回忆来回答问题,他们是在判断答案(例如:Burton & Blair, 1991)。例如,如果要求回答者回答在某段时间内他们去食品店购买食品的次数,他们通常是根据一般情况去估计,而不是把每个事情都回忆起来。这种观察结论导致研究者将设计策略转向帮助回答者作出更好的估计上。

最后,一个很重要的问题是我们应该认识到,有些事情虽然是研究者希望回答者能够回答,但却是回答者回答不了的。例如,如果我们真的想要了解医疗费用,我们就需要直接从费用的提供者或保险公司那里直接收集有关数据,以对回答人可以报告的数据(出自他们自己的开销)作出补充。

社会期望

在访谈中，有些事实或事件是回答者不愿意准确报告的。那些不合社会期望的健康情况，诸如神经病和性病等比其他健康情况被显著低报。对那些具有生命危险的住院经历，或者可能会给他们带来耻辱，或者使他们的生命受到威胁的经历，报告率较低（Cannell, Marquis, & Laurent, 1977）。有关酒精消耗总数的估计值存在严重的低报问题。尽管这种低报问题的确与回忆的难度有关，但也肯定与回答者对酗酒的社会道德心存顾忌有关。研究者发现，被逮捕和破产是另一种经常被少报而不可能忘记的事实（Locander, Sudman, & Bradburn, 1976）。

标准的访谈步骤可能会对人们报告的内容有一定限制。如果研究者希望有人承认一些很尴尬或非法的事情，那么就需要特别努力去劝说回答者，告诉他们所冒的风险极小和这个材料的重要性。在问到特别敏感的问题时，研究者应该考虑采用以下几个步骤（也可参见：Catania, Gibson, Chitwood, & Coates, 1990；Sudman & Bradburn, 1982）：

1. **使主观判断最小化，使精确性最大化。**重要的是注意引导语和问题措辞使用的词汇，以免令被调查者感到研究者对某种答案持否定态度。

在研究者与回答者进行交谈的过程中，研究者应该始终保持一种客观意识。如果有访谈员，那么他提出的问题和举止行为都会构成一种信息，回答者将据之推断研究者希望给出的答案。因此，研究者要特别注意回答者受到的语境提示对他们答案的影响。

2. **使用自填问卷的数据收集程序。**虽然数据不是结论性的，但是有一些证据表明电话访谈比个人访谈更易于产生社会意愿偏倚（Aquilino, 1994；de Leeuw & van de Zouwen, 1988；Fowler, Roman, & Di, 1988；Henson, Roth, & Cannell, 1977；Mangione, Hingson, & Barret, 1982）。在某些选

项上，让回答者自填问卷的形式比由访谈员提问产生的社会意愿偏倚要小（例如：Aquilino，1994，1998；Aquilino & Losciuto，1990；Dillman & Tarnai，1991；Fowler，Roman，Di，1998；Hochstim，1967）。对于那些进行敏感话题的调查来说，应该考虑用邮件调查或者集体填答。个人访谈调查中也可以包括一些自填问卷形式的问题：只要让回答者在一个小册子里回答一系列的问题就可以了。如果调查使用了电脑辅助，回答者则直接将他们的答案输入电脑，也可以取得同样的效果。例如，使用这样一种路数的调查报告显示，近来非法使用毒品的问题正日趋严重（Penne，Lessler，Beiler，& Caspar，1998；Tourangeau & Smith，1998）。最后，特纳、福塞斯和奥雷利（Turner，Forsyth，and O'Reilly，1998）指出，如果采用电话按键将答案直接输入电脑的办法取代访谈员访谈的方式，那么电话调查得到的与性和毒品有关的社会敏感活动的估计值，精确度更高。

3. **机密性和匿名性**。几乎所有的调查都向回答者保证处理他们的答案时坚持保密原则，非研究人员永远不会知道答案是由哪位回答者给出的。这些保证通常是在调查介绍中或者预达的信件中做出的。这些保证可以通过研究者签署的承诺书予以强化。没有标识码的自填问卷提供了一种不只是保密，而且也是匿名的方法。最后，如果调查是关于特别敏感的或涉及个人隐私的问题时，有一些详尽的调查策略，如随机回答技术（random response techniques），这些策略保证了不会将回答者与他们提供的答案联系在一起（参见 Fox & Tracy，1986；Fowler，1995）。

另外，调查研究的一个局限值得重视，那就是回答者愿意告诉研究者的内容乃是在研究者设计的资料收集条件下的。如果没有特别的努力，有些问题可能不宜在概率样本中问及。但是，本节讨论过一些程序，诸如努力为建立中立的答案语境以及强调资料收集过程的精确性和客观性等，对于大多数非敏

感性问题来讲，可能都是很有用的。任何问题，不管它看上去是多么好，都有可能使样本中的某些人感到尴尬。最好是在调查工具设计的各个阶段都保持对这一问题的敏感，以尽量减少社会意愿和窘迫感对人们可能给出的答案造成的影响。

提高描述主观状态的答案的效度

如前所述，主观性问题的效度与客观性问题的效度有不同的意思。并不存在客观的评价标准，评估一个主观性测量的效度，只有通过考察答案在期望的方面与其他有关问题的答案，或者个体的其他应该与之关联的特点的关联程度来实现（参见 Turner & Martin，1984，其中有关于影响主观性测量效度问题的深入探讨）。

基本上只有三种措施可以提高主观性测量的效度：

- 尽可能地使问题具有可靠性。回顾前面有关问题的信度、处理措辞的模糊性、使陈述标准化、答案形式的模糊性，以及尽量使问题对于所有的回答者都意味着同样的意义等问题的讨论。主观性测量即使只在一定程度上不太可靠，也会使效度有所降低。定序量表的信度问题有着特殊的意义，因为在主观状态的测量中，定序测量居于支配地位。提供的答案选项必须是一维的（例如：只涉及一个问题）和单调的（按顺序排列，没有倒置）。

有问题的量表

例 5.20　你是怎样评价你的工作的——非常有意义、有意义但有压力、不是很有意义但是没压力或一点意义都没有？

例 5.21　你是怎样评价你的工作的——非常有意义、有些意义、有意义或一点都没有意义？

问题5.20测量了两个特征即“有意义”和“压力”,它们不应该被列在一起,且未将所有的选项列出来。问题5.20如果既要问工作是否有意义又要问是否有压力,它就应该分成两个问题。在5.21中,可能有人会认为“有意义”(Rewarding)比“有些意义”(Somewhat rewarding)的意思更积极,因而对为什么要对类别进行这样排序感到不解。这两个问题都是比较常见的,因而也是需要尽力避免的。

- 如果沿一个连续统把人们按顺序分成若干等,那么类别多可能比类别少要好。但是,并不是类别越多越好,而是有一定的限度的。类别的数目应该以回答者在对答案进行排序时能够区分的精度为限。当类别的数量超出了回答者所能对自己的感觉作出区别的能力时,纵有很多的类别,也只会产生不可靠的噪声。并且,类别数量过多也使得问题难以操作,尤其在电话访谈中。但是,从能测量出回答者之间的变差的角度来说,更多的类别意味着效度的增加。
- 用测量同一主观状态的不同形式的问题,提出多个问题,然后把答案合并在一起构成一个量表,所有问题的答案,既受到被测量的主观状态的影响,又受到回答者或问题特点的影响。有些回答者拒绝极端的分类,有的人倾向于说同意而较少说不同意。多个问题甚至还能帮助回答者给出独特的答案,从而提高测量过程的效度(Cronbach,1951;DeVellis,1991)。

有关主观测量意义的最重要一点是它们的相对性。只有在刺激情况相同的时候才可进行分布状况的比较。在问题表述上的细微变化,答案选项数目的改变,甚至问卷中提问位置的改变,都会导致答案的重大差异(参见:Schuman & Presser,1981;Sudman & Bradburn,1982;Turner & Martin,1984,其中有很多影响答案分布因素的例子)。不能对主观性问题答案的分布状况直接进行解释,只有将不同样本回答同一问题的答案的

差异进行比较,或对答案关联模式进行研究之后,对答案分布的解释才有意义。

问题设计和误差

社会调查有一个明确的性质,那就是用问题的答案做量度的工具。很明显,答案作为一个测量工具,它的好坏程度是衡量调查估计值的质量的重要方面。问题可能会由于没有信度(产生不稳定的结果)或因为得到的估计值与真值之间在某一方向上存在某种一致性的误差而未能成为一个好的量度,(例如,低报酒后驾车被拘留的情况)。我们在怎样增加问题的可信度上所知甚少。本章所述的增加信度的方法可能比较好。尽管还可能加入一些其他的步骤,但是编制清楚的问题以便给不同回答者提供一致的测量,这对一个好的测量来说总是一个具有建设性的步骤。

效度的问题更复杂。在某种意义上来说,每个被测量的变量都需要用专门的研究来确定对它进行测量的一组最好的问题,并进而求出由此得到的量度的有效程度的估计值。本章所提到的很多有关改进报告的建议都来自一个长达 20 年的项目,这个项目是为了评估和提高与健康有关的变量的测量方法的(Cannell, Marquis, & Laurent, 1977; Cannell, Oksenberg, & Converse,1977)。有很多领域需要更多有关效度的研究工作。

通过更好的设计问题来减少测量产生的误差是成本最低的提高调查估计值的质量的方法。对任何调查而言,关注问题设计和先行测试(在第 6 章将谈到),并充分利用已有的关于如何测量要测量的内容的研究文献,这都是十分重要的。另外,进一步形成一部评估测量效度并予以报告的文献是十分必要的。除了大量的工作尚待完成之外,鲁宾逊、谢弗、赖茨曼(Robinson,Shaver ,and Wrightsman,1997)、麦克道尔与纽厄尔(McDowell,and Newell,1996)编辑了许多关于常用的多项选择

测量效度的资料。这些资料在告诉我们这些量度在多大程度上有效的同时,也指出它们在多大程度上还有待改进。

练习

使用本章讨论的标准来评估以下的问题是否是一个可靠的、可解释的和可用于分析的量度,如果可能,试行编写更好的问题。

a. 测量收入:你赚多少钱?

b. 测量健康状况:你的健康状况如何?

c. 测量对生活的满意度:你是怎样评价你的生活的——非常好、比一般水平好、一般、勉强可以或者非常糟?

d. 测量对堕胎法的看法:请告诉我你是否同意以下的陈述:堕胎确实是一个严重的问题;除非在紧急情况下,堕胎应该是非法的?

编写一个关于两个变量的可能关系的假设(例如,良好的健康状况与高质量的保健有关,或者高质量的住房条件与高收入有关)。描述你所需的资料,以便就这两个变量中的每一个给每一个人配值。然后起草一个问题(或一系列问题),以求它能在你在假设中具体指定的两个变量上对回答者的特征作出描述。问题的答案需为你提供所需的信息。指明你的问题要求得的是事实的还是主观的信息,明确得到的数据是定类、定序、定距还是定比的。

第6章　评估调查问题和调查工具

设计一个令人满意的调查工具包括几个方面的工作：选择能够满足研究目的的问题；对问题进行测试以确保它们都能按计划提问和回答；将问题制成能最大限度地方便被调查人和调查员使用的形式。本章将介绍设计令人满意的调查工具的几个步骤。

每一个调查都要有访谈计划或者调查问卷,前者是为访谈员编写的脚本,后者由被调查者自己阅读并填写。这些文件无论是纸质书面形式的,还是计算机程序形式的一般都可以看作是一种调查工具。

通过第 5 章的讨论,弄清了什么样的问题才是一个好问题以及怎样把问题作为测量工具。这些无疑都是设计令人满意的调查工具的基础。但是,为了设计一个好的数据收集的工具。我们还需要采取一系列切实可行的步骤。萨德曼和布拉德伯恩(Sudman & Bradburn,1982)、康弗斯和普雷瑟(Converse & Presser,1986)、布拉德伯恩和萨德曼(Bradburn & Sudman,1992)以及福勒(Fowler,1995)对此做了非常详细具体的探讨。

设计调查工具包括两个方面:确定要测量什么、设计并测试那些将要成为一个好的量度的问题。第一步通常是界定调查目标,当然这些目标可以根据随后的问题测试进一步修订。然后就是选择和测试问题。调查工具开发过程包括以下几个步骤:

- 讨论焦点组。
- 草拟一系列尝试性的问题。
- 批判性地复查问题以找出不足。
- 单独的实验室访谈(不要搬用建议的数据收集程序)。
- 把问题组合成调查工具。
- 运用近似于建议的数据收集程序进行前测。

界定调查目标

设计一个令人满意的调查工具,其先决条件是确定要测量的内容。这看上去很简单且不证自明,但这一步骤经常被忽视,从而影响调查结果。第一步,也是很重要的一步是写出调查预期达到的目标。在设计调查工具时,研究者常抵挡不住诱

惑,要添加一些相关但于研究目的无益的问题。抵制这种诱惑的方法,是对调查目的作出明确的表述,并以此保证对某一特定调查领域的调查是能够测量的。其次,研究者应当列出为达成研究目的所要测量的内容的清单。内容清单所列出的不是问题,而是要测量的变量。这些变量要以有意义的类别或方面的形式列出。

除了要列出测量的变量之外,还应制订一个分析计划。假设,在样本设计上我们将会有一个良好的开端。研究者也将对总体中哪些子群体需要特殊的估计值这一问题作出明确的决定。然而,现在研究者仍需对这些观点作进一步的提炼,以便列出一张明确的清单:①把哪些变量设计为因变量,使用什么样的集中趋势量度(如均值或值的分布);②为了了解关联的分布和模式,我们需要哪些自变量;③为了解释观察的模式和检验竞争假设,我们需要把哪些变量作为控制或干扰变量。

总之,目的陈述书、测量变量的分类清单和分析计划草稿这三份文件是开发一个调查工具的基本组成部分。

初步的问题设计步骤

焦点组(focus group)

在草拟一组结构化的问题之前,有必要就所要研究的问题与研究总体中的人进行焦点讨论。讨论的主要目的是对回答者回答问题的实际情况与嵌入在研究目标中的抽象概念进行比较。

例如,研究目的是测量看医生的次数,群组讨论就可以集中在怎样才算是看医生这一问题上。这里两个关键概念是“看”(visit)和“医生”(doctor)。可以询问参与者与医生的各种交往(如电话咨询、照 X 光或实验室测试、接种疫苗),并问他们是否认为这些就是“看医生”。也可以询问他们所接触的

关涉他们健康的各种人员（如心理学者、精神病医师、医生助理、眼科医师、验光员、临床医学家）的情况，以及是否认为这些人也是医生。

仅仅这个讨论就可提供给我们至少三种信息：

- 可以算作是“看医生”的交往类型。这一信息可以帮助研究者改进研究目的和提问的措辞，以明确应当或不应当包括某些内容。例如，要不要包括电话咨询呢？如果医生办公室有见习护士，算不算医生呢？
- 人们知道些什么。例如：是不是每个人都清楚精神病医师有医学博士学位而心理学家却没有呢？关于医疗保健工作者的背景、培训和资格，人们是怎样认识和理解的呢？
- 对一些关键词语或术语的理解。“医生”是指获得医学博士的人呢？还是指更普通的身着白大褂、提供健康服务的专业人员一类的人呢？

参加一次焦点组讨论的人数最好在 6 ~ 8 人，讨论的内容最好围绕人们对调查要测量的内容的理解、经验和感受等。小组讨论人数到底多少为宜，可能尚无定论，要看具体情况而定。但是，在调查工具开发的初期进行几次焦点小组讨论，将最终有益于调查工具的制定则是确定无疑的。

草拟问题

拥有了所要测量内容的清单，研究者就尝试去寻找一个或一系列问题来构建测量清单上那些变量的工具。诸如有关背景或人口统计之类的很多问题，对很多调查来说都是常规性的问题。审视一下芝加哥大学（University of Chicago）民意研究中心做的“一般社会调查（General Social Survey）”问卷是有裨益的。另外，还可以通过密歇根大学的“高校间政治和社会研究协会”（ICPSR）在线获得很多调查的资料。各主要调查机构的调查工具的原件复印件也是有用的参考资料。研究者可以从这些资料中获悉如何表述详细而精确的问题，如何生成标准化

问题，以及如何格式化调查工具。

借鉴他人已做的工作是非常明智的。当然，最好能回顾一下那些以前做过这个研究课题的研究者用过的问题。另外，如果一些问题已经在其他样本中使用过了，那么收集可比的数据就能增加研究的普遍性。但是，有人曾经用过某个问题这一事实并不能保证它就是一个好问题，或者就是一个适合于某个给定调查的问题。由于研究者不加批判地使用，导致了很多有缺陷的问题被反复使用。对所有问题都应当加以测试，以确保它们适用于特定的研究总体、背景和目标。

前调查评估(presurvey evaluation)

批判性的系统复查

一旦一系列问题草拟完毕，紧接着就是对它们进行批判性的系统复查。莱斯勒和福塞斯(Lessler & Forsyth，1996)列出了一张对拟就的问题进行复查时应该注意的问题的清单。福勒(Fowler，1995)也列出了一张对拟就的尚未应用于试调查的一组问题进行复查的规范的清单。尽管这两份清单都不能算是很全面的，但我们还是能用它们来识别问题的某些特征，找出拟就的问题中存在的某些问题。这两种清单中的任何一种，都可以帮助我们找出需要进一步修订的问题。不仅如此，我们也可以用它来标出那些在下一步应该予以关注的问题。

认知实验室访谈

一旦问题草拟完毕、经过批判性地复查及修改之后，下一步就是确认它们是不是人们都能一致理解并作出回答的问题。焦点组讨论虽然能在人们对问题的理解方面使研究者有更深入的了解，但在问题的具体的措辞是否适当，或问题回答的难易程度的判定上却不能给研究者提供什么帮助。在设计问题

的早期阶段，研究者也可以通过朋友、亲戚和同事来试验设计的问题，从中也可以有所收获。大多数调查工具的早期版本都包含一些含糊不清的问题，这些问题读起来难懂，实际上任何人都无法回答。

问题拟定好后，在进行正式实地试测之前，还要进行一次更为正式的测试，这就是人们通常所说的认知测试（DeMaio & Rothgeb，1996；Forsyth & Lessler，1992；Fowler，1995；Lessler & Tourangeau，1989；Willis，DeMaio，& Harris Kojetin，1999）。虽然认知访谈形式各异，但它们确实有一些共同特点。首先，接受访谈的是一些志愿者，为了帮助研究者理解问题是如何起作用的，他们花费的时间比资料收集本身实际所需时间多得多。通常情况下，受访者都有报酬，置身于实验室环境，整个访谈可以进行录像。

这些访谈通常不是一般访谈员能做的。有时是由认知心理学家来做访谈员，有时是由调查者本人或高级访谈督导员来做的。在大多数情况下，访谈员需要对每个问题的目标都十分清楚，以便能觉察出受访者在问题理解或答案形成过程中出现的问题。一份典型性的草案旨在向受访者提出一组准备好的问题，然后以某种方式收集有关他们对问题的理解和回答问题方式方面的信息。有时，在受访者准备答案时，要求他们"大声地想"（think aloud）；有时会问受访者一些有关他们怎样理解及回答问题方式的问题。最常见的两种任务是：

- 要求受访者用自己的话说出他们是如何看待所提问题的。
- 要求受访者解释他们为什么选择某一特定的答案，而没有选择其他答案。

关键是要收集足够的有关受访者对问题的理解和答案准备过程中的信息，以评估他们是否是以研究者所希望的方式来完成这一任务的。大多数认知测试要回答以下四种问题：

- 对问题的理解是否一致？
- 受访者是否拥有回答问题所需的信息？

- 答案是否准确地描述了受访者必须诉说的东西？
- 答案是否能有效测量设计要求的问题要测量的东西？

从实验室访谈所得的数据是比较有限的。通常这样的访谈只能做为数不多的几次（一般不到 10 次），因为这种工作十分耗费人力，而在大多数机构中，只有很少一部分人能主持这样的工作。其次，访谈是在人为的环境中进行的，一个志愿者能够并乐意执行的任务，可能在一个截面样本（a cross-section sample）中是无法执行的。尽管如此，人们还是越来越认为这种访谈是设计和评价调查工具的关键一步。那些在实验室环境中都无法一致理解或回答的问题，在用于实际调查时，其效果肯定不会太好（Royston，1989）。在实地前测中要确切地确定在回答问题过程中存在的理解问题和困难之处，不如在实验室认知访谈中那么容易，因为后者可以对这些问题进行实时的考察。

实验室认知访谈法大多用来测试访谈草案，但理解和答问方面的问题与困难同样存在于自填问卷的调查中。尽管自填答案式问题的标准测试形式，正如我们下面将要讨论的那样，也常常使用一些与认知访谈中十分相似的指示性问题，但是回答人在理解方面的问题毕竟在口头进行的问答中表现得更加明显。因此，在对为自填问卷式访谈设计的问题进行测试时，使用一个口头认知访谈可以有效识别那些标准前测所不能识别的问题。

调查工具的设计、格式和编排

一旦问题准备就绪并将付诸最终的前测时，必须对它们进行编排，以便于访谈员或自我填答使用。第一步就是给问题排序。很多研究者都喜欢将相对比较简单、直截了当的问题放在最前面，以吸引受访者。通常，把需要认真思考的或人们认为

很敏感的话题放在调查工具的中间或后面。一个实用的好办法就是按节给问题编号：A1、A2、B1、B2 等，这样，增加或删除问题后就不必再为每个问题编号了。

不管调查是采用访谈员访谈还是自我填答的方式，问卷的编排和格式化都是为了尽可能地便利访谈员或回答者。对于由访谈员操作的调查工具，以下规则有助于达成这一目的：

- 在字体编排要区分向回答人宣读的用词和用于引导调查员的用词时，约定俗成的做法是引导语用黑体，需要宣读的话语应用普通的字体。*
- 如果访谈采用纸笔方式而没有电脑辅助，对那特定访谈对象可忽略不答的问题在指导语中做出明示。当然，电脑辅助的访谈可以自动跳题。
- 用括号表示可供选择的词语，可采用诸如“（他/她）”或“（丈夫/妻子）”这样的习惯用法。如果访谈员看到括号内的提示，处理起来就比较容易了。另外一个习惯做法就是，在遇有访谈员必须提供，而在问题中又没有出现的词语时，该词可用黑体（如，配偶）。电脑辅助的访谈能提供供选择的词语，而不是由访谈员自己根据情况调整措辞。
- 进行仔细地检查，以确保需要访谈员说的话的确都写下来了。要检查的内容不仅包括叙述问题的语汇，还包括过渡词、问题的介绍、必需的定义和解释。

这些基本原则也同样适用于自我填答问卷式调查，即主要目的是使问卷用起来更便利。如果有什么区别的话，那就是自我填答问卷调查的格式更重要。与访谈员相反，受访者没经过训练，他们本来就缺乏做好问卷的动机，加之受访者并非是按处理问卷的能力来选取的，所以格式问题就显得更加重要。自我填答问卷调查有如下五个指导原则：

---

* 这里根据中国情况，对引文略加修改，把原文中的“大写字母”改为“黑体”，“小写字母”改为“变通的字体”（校者注）

- 在多数情况下，自填式问卷是不言自明的。通常不需要有读卷须知之类的引导语的，因为即使有这样的引导语，回答者的理解也未必全都一致。
- 在多数情况下，自我填答问卷仅限于使用封闭式答案。受访者唯一要做的事是在方框中打勾，在答案上按鼠标，或在数字上画圈。如果要求受访者用自己的话回答问题，答案常常是不完整、模糊而难以编码的，因而作为量度，它们的价值是极其有限的。
- 在自我填答问卷中问题使用的形式不宜过多。在一个工具中，如果受访者完成同种类型的任务或回答同种类型的问题越多，他们感到迷惑不解的可能就越少，因而他们做起来也就越容易。
- 问卷的版面应该清晰、整洁。版面过于紧凑（或其他在一页上排很多问题的做法），与把同样多的问题编排在更多页面上，使页面有更多空白相比，会使调查的实际回答率有所降低。
- 通过书面和可视线索为受访者提供更多信息，从不同的角度向回答者传递相同的信息——如何继续。如果人们有可能不清楚究竟要求自己做什么，那么他们可能不清楚究竟应该怎么做。应该尽力使每一个问题都变得简单明白。

电脑辅助测量工具的设计还在继续发展中。毫无疑问，上述很多原则都同样适用于电脑辅助的测量工具。但是，库珀、汉森和萨达斯基（Couper, Hansen, & Sadowsky, 1997）发现访谈员在使用一些复杂的 CAPI 程序时难度很大，这说明我们还需要在这方面做更多的工作（同样可参见 Couper, 1999）。迪尔曼（Dillman, 2000）介绍了他在设计最佳互联网调查工具方面取得的成果，这项工作同样也正在进展之中。

实地前测

一旦调查工具设计完成,研究者认为基本上可以使用了,这时就应该对调查工具和调查程序进行实地前测。进行这种前测的目的是找出数据收集草案和调查工具如何在真实环境下运行的。

访谈计划的前测

由权威调查机构做过的传统的前测,其通常做法是,从与要调查的总体相同或者相似的总体中抽取受访者,由经验丰富的访谈员对这些受访者做20~50个访谈。访谈员在这种前测中扮演两种角色:一方面,他们是访谈员,执行访谈任务;另一方面,他们又是数据收集过程的观察者,要向研究者报告调查程序和调查工具中任何有待改进之处。尽管有时访谈员的汇报有可能是个别进行的,但一般诸如这样的反馈可能都是在集体汇报会上进行的。

上述前测是调查设计过程的重要组成部分,它特别重要的一个功能就是从访谈员的角度来测验测试工具的可用性,包括问题和版式。但是,这种测试也有一些局限性。访谈员用以判定存在问题的标准通常不很明确,并且几乎可以肯定地说,访谈员们在认为什么算是问题上也观点不一。另外,群组讨论在收集有关前测经验的系统性信息方面也不是一个完美的办法。

为使前测经验更加系统和更有价值,研究者补充了一些步骤。创举之一就是要求访谈员除了群组汇报之外,还要给每个问题填写一个简明的分级评分表格,在表格中要求访谈员评价每个问题是不是:①措辞是否便于宣读;②受访者理解问题的一致性;③受访者能否准确回答问题(Fowler,1995)。很明显,访谈员不得不猜测受访者是否理解了问题并能准确回答,但是,无论如何他们都会这样做。表格的优点在于,它要求访谈

员除了关注惯常注意的调查工具的实用性之外，还要全面系统关注问题设计的上述方面。此外，让访谈员对这些问题分级，可以让主持调查的人对访谈员的报告加以概括，用更一致的方式识别设计的问题中存在的不足。

就实地前测而言，一种更重要且更有用的创新，是使用磁带录音和行为编码来评价调查问题。如果受访者同意（这几乎不成问题），就可轻而易举地录下前测访谈，不管它是面对面还是通过电话进行的。然后，训练有素的编码者通过听录音并用一致的方式对问答过程中出现的问题作出评价。

业已证明，在确定调查问题中存在的问题时，有三种行为是最为重要的（Fowler & Cannell，1996；Okesenberg，Cannell，& Kalton，1991）：①访谈员是否按照书面文字读出了问题；②受访者是否要求解释说明；③受访者最初给出的答复是不是不够准确，需要访谈员进行探询。我们发现，在访谈中，调查的问题要么总是引起这类行为的发生，要么不会引起这类行为的发生，也就是说，有些问题访谈员总是在误读，致使受访者要求进行澄清，或者总是不能以准确的方式回答问题。对这样的行为编码固然不能识别所有那些受访者未能一致理解的问题。然而，我们已经有证据证明，在前测访谈中，这三种行为中的任何一种行为的发生率为 15% 及以上时，这一问题或很有可能会产生歪曲的数据，或明显受到访谈员的影响（Fowler，1991；Fowler & Mangione，1990）。

前测访谈中行为编码的另外一个好处是其结果具有系统性，而且可以复制。因此这样进行的问题评估，其过程不仅不会受研究者和访谈员的主观意见影响，而且得到的有关不恰当的问题的评估证据是具体的和可重复的。

从计算机辅助访谈的前测中可以得到跟踪文件（Trace files），这是第三种信息来源。当访谈有计算机辅助时，可以重新得到访谈员敲击的关键键。这些文件可以识别访谈员在何处返回到前面的屏幕或前面的问题。返回到前面的问题会减缓访谈进程，也通常表明了那个问题的流畅性没有设计好。查

看“帮助”功能的使用情况，就可以确定什么地方需要帮助以及各种帮助功能到底有什么作用。此外，检查跟踪文件还有一个好处，那就是检查得到的结果是系统的和可以量化的（Couper，Hansen，& Sadowsky，1997）。

自填式问卷的前测

正相反，与访谈调查工具相比，自填问卷调查工具需要做更多的前测工作，原因是前者中的访谈员可以解决一些研究者在设计调查工具时没有解决的问题，而后者却不能。遗憾的是，对自填测量工具进行前测并不是一件容易的事，因为理解上的问题和回答问题时的困难并非显而易见的。尽管有人采用观察人们如何填表，或如何与计算机互动这样的方法来识别模糊的或令人迷惑的问题和指导语，但这些方法不如录音和访谈的行为编码那样令人满意。

或许对自填问卷进行前测的最好方法就是亲自与一组可能的受访者交流。如果是基于电脑的调查，受访者可以使用便携式电脑。首先，受访者应像自己是调查的一部分一样完成调查问卷，然后研究者可以组织进行关于调查工具的讨论。第一个话题当然是指导语是否明确，第二个话题是问题是否明晰，第三个话题是在对什么是期望的答案的理解上，和按规定的要求给出问题的答案这两个问题上是否有困难。

除了团体测试之外，基于计算机的调查工具的可用性往往还得益于一对一的测试。在这种测试中，观察受访者与计算机及问题间的互动。

调试计算机辅助调查工具

让访谈员或受访者测试调查工具，这可以提供关于使用是否方便的信息，但不能提供资料收集草案是否正确的信息。需要关注的重点是“跳过（Skip）”指令。

计算机辅助的一个巨大优势是帮助受访者和访谈员正确定位关联事件（contingency）：当哪些问题被问及或者这些问题

该如何措辞与前面的问题给出的答案相关时，要保证“跳过”指令的准确性就要对问卷调查工具的最后版本进行仔细校对。但是，检查计算机辅助调查工具的准确性所面临的挑战远远大于检查问卷调查工具。其难点在于测试者看不出跳过了哪些问题，因此他们可能忽略某一问题本应问及却被略过这一事实。校对打印出来的程序并做大范围的测试是很重要的步骤。但是，如果调查工具很长而且还包含复杂的关联事件，仅做这些就不够了。规范的做法是，调查一旦开始，我们就应该把早期回收的答案的分布制成表格，对之进行必要的检查。因为只有通过这样的检查，研究者才能确定关于关联事件的指令是否按要求运行。

调查工具的长度

一个高质量的前测成果之一是可以帮助我们确定完成一个调查工具需要的时间。判定访谈长度的标准应该包括成本、回答速率的影响以及受访者的能力限制和答题意愿等方面。自填问卷的长度对成本和回答速率的影响，随调查人群和调查主题而异，不能一概而论。另外，受访者接受访谈时间究竟能有多久，这一问题也难有定论。

如果研究者发现要问的问题多于能问的问题，那么有两种选择：当然最简单的办法是，研究者只要删除一些问题就是了；备择的办法是，某一组问题，研究者只问被调查人中一个有代表性的子样本。这种方法不仅会增加调查的复杂性，而且也会降低这些变量的估计值的精确度。但与将问题完全删除相比，这种方法还是比较可取的。计算机辅助资料收集的一个明显优势就是，借助它可以方便地完成这样的设计。

总　结

人们曾经认为,对问题的评价在很大程度上是一个主观过程,结果取决于访谈员和研究者个人的品味和偏好。而现在我们已经知道,我们对问题做的评估可以超越这样的主观性。对所有接受访谈的人来说,调查问题的意义都应该是一样的。而回答调查问题则应当是绝大多数或几乎全部接受访谈的人都能做到的一件事。为了做好一个访谈,计划在访谈过程中使用的话语应该以文字形式正确地表达出来,以便访谈员能照本宣科进行操作。

显然,无论问题表述得多么清楚,总有些接受访谈的人会遇到困难,而且总有些访谈员会误读。在问题不得不被修改以前,我们须判断这个问题有多少不足之处。设计和评价调查工具的关键是收集有关理解、问答以及访谈员和受访者如何使用草案的信息。这样就能判断是否需要修改调查问题和调查工具。在调查真正实施之前,对调查问题和调查工具作出充分的评价,是使调查获得预期效果的关键。这是减少调查估计值的误差的最经济的方法之一。尽管在确定什么才是问题评估最有效和最经济的方法这一问题上,我们还有许多工作要做,但前面介绍的那些方法已经都是很有用处的。如果能把它们付诸实施,定将对提高调查资料质量产生重大而积极的影响。

练　习

运用第 5 章练习中设计的问题,把它们转换成访谈员能以标准化方式控制的一组问题。根据需要进行前测并修正。再把同样的问题转换成自填问卷调查形式,并进行前测。

第7章　调查访谈

访谈员从三个方面影响调查估计值:他们对调查能达到的回答率起重大作用;他负责培训和激励受访者;他们是访谈互动和以标准化的无偏的方式进行的问答过程中重要的一方。本章讨论的问题包括调查员的选择、培训和督导,以及研究者给予访谈员的一系列程序对最大限度地减少在调查中与访谈有关的误差的意义。

访谈员工作概述

尽管有些调查是用自填法进行的，但无疑使用访谈员，包括面对面访谈和电话访谈，来提问并记录答案的方法仍是诸多调查测量方法中较为常用的方法。由于访谈员在数据收集过程中作用重大，因此他们就极有可能对收集的数据的质量有很大的影响。访谈员的管理是一项十分困难的工作，对于采用个别访谈的研究来讲，情况尤其如此。本章的目标是让读者明白访谈员应该做什么、访谈员管理的恰当程序，以及访谈员的管理和行为对调查估计值的意义。

访谈员在调查数据收集过程中的作用可概括为以下三个方面：

- 选定受访者并谋取合作。
- 培训并激励受访者做好访谈工作。
- 提问、记录答案、探查不完整的答案以确保答案符合提问的目的。

获得合作

访谈员要与受访者取得联系以获得合作。这一工作的难度因样本不同而不同。要让受访者在他们愿意回答问题的时候，马上能够找到访谈员，访谈员要随时准备好与难以接触的受访者取得并保持联系，并在面对面访谈中，访谈员要能够并愿意去受访者所在地。

尽管很多样本个体乐意接受访谈，但要得到事先未通知到或最初不太愿参与访谈的人的合作，这是访谈员必须做的最困难的、却也是最重要的工作之一。访谈员在这个问题上比在任何其他问题上都更容易失败。

无疑有些访谈员比其他访谈员在争取受访者的合作方面

更加出色。同时可清楚地看到,不同的个人风格也会产生不同的影响。高效率的访谈员,有些办事很有条理,而有些则风度翩翩。经验告诉我们,那些在获取受访者合作方面很出色的访谈员有两种共同特点:第一,他们都非常自信。他们在陈述研究项目时似乎完全相信受访者肯定会合作。他们说话的语调和内容都表明他们相信访谈肯定会成功。第二,他们有迅速调动人的兴趣的诀窍,因此使他们与被调查人的交往很快就变得专注且具有个人的色彩。争取得到被调查人合作这一工作也许的确有着非常强烈的任务取向,但它何尝又不是关系着个人的需求、忧虑和处境呢?

虽然这些访谈员的技巧对所有调查都十分重要,但对于电话调查来讲这些技巧更是意味着一种挑战。因为在电话调查中受访者事先没有得到通知(如随机数码拨号访谈),或者对调查主题不感兴趣。

培训和激励受访者

受访者在访谈中的表现,诸如报告的精确性等,与他们对访谈的定向有关。业已证明,访谈员对受访者的目标设定有着重要作用(Cannell & Fowler,1964;Cannell,Oksenberg,& Converse,1977;Fowler & Mangione,1990)。例如,访谈工作中草率的访谈员会鼓励受访者快点回答问题。而一个有条不紊地读出问题的访谈员,则在用一种非言辞的方式告诉受访者,他愿意花一点时间得到经过认真考虑的、准确的答案,其结果是,他们果然得到了更为准确的答案。研究也表明,访谈员鼓励受访者的方式会影响受访者对自己应当做什么及做得如何的感知(Cannell,Groves,Magilvey et al.,1987;Cannell,Okesenberg,& Converse,1977;Fowler & Mangione,1990;Marquis,Cannell,& Laurent,1972)。

毫无疑问,大多数受访者都不太了解自己应当做什么和如何去做。访谈员明确或含蓄地教给受访者该怎样做,这一点常常得不到应有的重视,但它却是访谈员工作中十分重要的

部分。

做一个合乎规范的访谈员

调查研究者乐于假设,答案中的差异可归结于受访者必须要诉说的内容本身(即他们的观点和经历)存在的差异,而不是归结于他们所受刺激(即提问的措辞、提问的背景及提问的方式)的差异。绝大多数对访谈员的培训的目的都是使受训者成为一个不对他们得到的答案施加影响的规范的访谈员。研究者试图从五个方面对访谈员的行为加以规范:陈述研究和任务的方式;提问的方式;探讨不完整答案(即没有达到问题目的的答案)的方式;记录答案的方式;访谈中处理人际关系的方式。下面对这五个方面做一个简单的介绍:

1. **陈述研究**。受访者应对研究目的有一共同的理解,因为目的性与他们答题的方式有关。对诸如保密性、自愿性以及谁运用结果等问题的臆断都会影响到答案。优秀的访谈员会让所有受访者对访谈项目有相似的取向,这样就能使访谈背景保持一致。
2. **提问**。调查问题应当完全按其文字表达的形式宣读,不能有任何变动或措辞改变。有时,即使问题用词的细小改变都会对问题的答案产生重要的影响。
3. **探查**。如果受访者没有完整回答问题,访谈员必须提出一些追踪问题,以探出更好的答案,这就是探查。访谈员应当间接地探查不完整的问题,即尽量不让答案出现重合的情况。如果调查工具设计得好,包括重复提问在内的,诸如"还有什么吗?""能多告诉我一些吗?""你的意思是什么?"这样的标准化探查,基本上可以处理调查遇到的绝大多数不完整答案的问题。
4. **记录答案**。记录答案应标准化,以确保在这一阶段不会产生访谈员引起的变差。提出开放式问题时,访谈员要逐字记答案,也就是说要准确记录下受访者使用的词语,不能

做阐释、概括或者遗漏。在封闭式问题中，给了受访者选项，访谈员应该在受访者真正选定后才记下答案。如果访谈员将受访者的话编成他们并未选择的类别的码，那么就有可能产生不一致性。

5. **人际关系**。访谈中的人际关系也要以标准化方式处理。访谈员不可避免地会把某些人口统计学特征带入访谈，如性别、年龄及教育情况等。但是，通过强调互动关系的专业方面并将注意力集中到研究任务上，就能最大程度地减小关系中的个人方面。一般地，应教导访谈员不要讲述他们自己的事情或者表达与访谈主题内容有关的个人意见和看法。访谈员不能就受访者做出的答案有任何评价。总之，要避免访谈员个人的、特殊的行为特征的影响，因为这些特征在访谈员之间是不同的。表现得像专业人士而不像朋友，这有助于实现访谈员与受访者间关系的标准化。没有证据表明友善的人际关系风格会从本质上提高报告的准确度，而且它很可能对报告的精确度有负面的影响(Fowler & Mangione,1990)。

如果访谈员和受访者的社会背景不同，情况就复杂得多了。在这种情况下，交流就不如背景相似时那样随意和轻松。有证据表明，那些采取措施来缓冲这种复杂性(如交流时加入一些幽默)的访谈员可能得到更有效的访谈(Fowler & Mangione,1990)。但是，努力使受访者放松不应该偏离基本的专业化交往，目标应放在取得令人满意的结果上。

访谈员工作的意义

现在我们可以清楚地看到，访谈是一项困难的工作。此外，这一工作做不好就会导致三种不同的调查数据误差：

- 如果访谈员不努力取得受访者的合作，样本就会没有信度，从而产生偏倚。
- 如果访谈员影响数据的方式不一致，而在这种不一致严重

到一定程度时，调查估计值的准确性就会降低，估计值的误差就会加大。

- 如果访谈员不能正确训练和激励受访者，或未能建立报告所需的恰当的人际关系，这样的情况严重到一定程度，那么答案就会系统地不正确或有偏倚。

由于这些情况都可能引起误差，所以，应该提倡研究人员使用优秀的访谈员。很多方面都会影响访谈员的工作质量：征募和选拔、培训、监督、设计令人满意的问题以及运用有效的程序。在下面五节，我们将讨论上述这些方面对访谈员工作可能的影响。

访谈员的征募和选拔

访谈员条例中规定的要求访谈员具备的某些特征与调查数据的质量本身并没有什么关系：

- 访谈员要具备良好的读写技能。即使不是大多数，至少许多访谈员都要在电脑上工作，因此访谈员还要会打字并基本熟悉电脑。绝大多数调查研究机构要求访谈员有中学毕业文凭，而很多机构则都要求或乐于使用有大学学历的访谈员。
- 访谈大多是兼职工作，很难做到每周工作 40 小时来做基本人口调查，调查机构对访谈员的需求总是时多时少。因此，可能做访谈员的人通常是可以接受不连续性的收入，或处于两个更稳定工作之间的人。访谈员的报酬对于一个大学毕业生来说通常是不算高的，而且经常没有福利，如健康保险等。访谈调查人员很难仅靠这份工作的收入来维持长久的生活。
- 个人家庭访谈员在时间上必须有弹性，调查要求访谈员在受访者有空时就到场。电话访谈的优点之一就是可事先

为单独的访谈员安排好工作的班次，尽管对所有的一般总体的调查几乎都安排在晚上和周末。

- 进行入户访谈的访谈员必须随时都可以行动，残疾人或者没有汽车的人不宜承担这样的工作。但电话调查的访谈员则没有这些限制。

除了这些实际工作的要求以外，我们可赖以对访谈员的优劣做出的判断，来自研究的根据不多。例如，经验丰富的访谈员更容易得到受访者的合作，原因仅仅是因为那些难以获得受访者合作的人已经被淘汰了，但是，没有记录表明经验会对数据质量有积极影响。有迹象证明，随着时间的推移，访谈员会变得漫不经心，因而收集到的数据的质量也每况愈下（Bradburn，Sudman，& Associates，1979；Cannell，Marquis，& Laurent，1977；Fowler & Mangione，1990）。

与之类似，要求访谈员拥有研究主题方面的专业知识也是多此一举的。实际上，当受访者没有表达清楚的时候，知识渊博的访谈员会自以为是地认为受访者在说什么，他们会比没接受过这个领域训练的访谈员更多地把自己的意见加到受访者的回答中。因此，除非访谈观察或报告要求访谈员必须拥有渊博的知识背景，通常没有专业知识背景但受过训练的访谈员才是最佳人选。

尽管一些来自截面样本中的证据表明，女性访谈员可能比男性更好一些，但访谈员的年龄、教育和性别则与数据质量的关系不大（Fowler & Mangione，1990；Groves，1989）。一般讲，研究者总是选派最好的访谈员对受访者进行访谈，而不论其人口统计学的特征究竟如何。例外的情况发生在调查主题直接关系民族或宗教（或任何其他人口统计学特征），且涉及同一或不同群体的人的感受时。例如，如果要采访人们的泛犹太主义观时，访谈员是否为犹太裔便会导致答案的差异性（Robinson & Rhode，1946）。同样，黑人和白人也会因访谈员的肤色而对种族问题作出不同的回答（Schuman & Converse，1971）。

但是,要特别注意的是,种族匹配并不一定能提高报告的质量。有两项研究发现,黑人受访者接受白人访谈员关于福利收入(Weiss,1968)和投票(Anderson,Silver,& Abramson,1998)问题的访谈时,比接受黑人访谈员访谈时表达得更准确。

毫无疑问,研究者应该考虑调查主题与访谈员和受访者的人口统计学特征之间的互动关系。如果种族划分(或其他特点)对将给出的答案极为重要,那么就需要考虑控制访谈员与受访者特征之间的关系,以便测量访谈员对数据的影响(Groves,1989)。但对于大多数调查来说,实际操作诸多困难和控制调查员分配的成本,加之其影响的不可预测性,都不支持控制访谈员和受访谈员的人口统计学特征。

最后,志愿访谈人员几乎总是不能很好地完成概率抽样调查,失败原因有如下几个方面:第一,很难要求志愿者参加冗长的训练,因而他们常常得不到系统的训练;第二,很难拒绝那些出于自愿,但素质不高的访谈员,因而使用志愿调查员的调查的调查回答率通常都不太高;另外,志愿访谈员减员率通常都很高。

上述讨论为研究者在选择访谈员时提供了指导性原则。在某些非常特殊的情况下,访谈员的民族背景、年龄或者性别会影响答案。例如,十几岁的青少年在接受年纪大的女性访谈员的访谈时表现会有很大差异(Erilich & Riesman,1961)。但是,在绝大多数调查中,是由特定的研究工作需求支配着访谈员的组合。因为个人的背景或个性特征而排除某些人做访谈员,这是没有根据的。相反,要组建优秀的访谈员群体,关键是要给予他们良好的培训和精心的督导。另外,由于无法事先识别优秀的访谈员,事后裁减掉那些能力较差的访谈员,对于建立素质优良的访谈员队伍来说是至关重要的。

培训访谈员

有关访谈员培训方面的经验可谓不一而足，种类繁多。很明显，投入培训的时间长短、培训课程的类型及课程内容都取决于特定的组织环境以及访谈员将要做的工作。另外，对于在初始培训尤其是最初的实地经验培训、继续学习及开始工作后的再培训等方面投入多大的精力，人们并未达成共识。尽管如此，所有关注数据质量的专业调查机构对新招募的访谈员都会进行至少某些类型（通常是面对面）的培训。下面概括了恰当的访谈员培训应当包括的项目。

培训内容

培训内容既包括适用于所有调查工作的一般知识，也包括专门针对特定访谈工作的知识。要讨论的共同话题包括如下几个方面：

- 联系受访者和介绍研究的程序。
- 在设计调查工具时作出有关问题措辞和跳答的规定，以使访谈员能一致和规范地提问。
- 间接地探查不完整答案的程序。
- 记录开放式问题和封闭式问题答案的程序。
- 无偏地处理人际交往问题的原则和指南。
- 如何使用计算机辅助访谈程序。

此外，很多研究机构认为，让访谈员感受如何使访谈适合于整个研究过程是个不错的主意。为此，他们常试着让访谈员熟悉抽样过程、编码、分析的类型以及调查结果的报告。这类信息有助于访谈员回答受访者提出的问题，并且还可激励访谈员并帮助他们理解访谈工作。

具体到某一特定的课题，访谈员还应知道以下几点：

- 课题的具体目的，包括资助者、研究的总体目标以及预期的研究的使用情况。这些信息是为受访者提供正确的问题答案并取得他们合作的基础。
- 抽样的具体方法，它也将为回答受访者提出的问题提供根据。此外，还需要一些关于如何实施基本样本设计的培训。
- 详细介绍特定问题的具体目的——不必介绍它们在分析中的作用，但至少要介绍它们将会提供什么样的信息。
- 关于保密以及给予受访者各种恰当承诺的具体步骤。

培训的步骤

培训访谈员有六种基本方法：书面资料、讲座和报告、基于电脑的培训教程、有计划的练习、角色扮演及对最初的访谈进行观察。书面资料通常有两类，首先，最好有一本详尽的访谈员手册，全面描述访谈过程；另外，对于每一特定的研究，都应该有一些针对特定项目的说明。在对访谈员当面进行培训和项目在当地实施时，有望节省书面材料的准备工作。不过新培训的访谈员都认为，培训期间的资料和信息太多，难于马上吸收。因此把程序写成书面材料，能使访谈员有更充裕的时间进行复习，因而能更清楚和准确地理解陈述的内容。

显然，不管是只培训一个还是培训一群访谈员，讲座和展示在任何访谈员培训中都起作用。除了总体展示必需的程序和技能外，大多数培训者都发现，展示标准化访谈是让访谈员明白怎样开展访谈的快捷而有效的方式。讲座大多使用录像带作补充，见习访谈或其他访谈活动的录像带是培训访谈员的有效工具。还可以购买普通培训磁带用来培训访谈员。

计算机辅助访谈的广泛使用意味着在访谈员培训过程中必须向访谈员教授如何使用以电脑为基础的测量工具。使用最为广泛的调查系统都有可以编入一般访谈培训材料中去的基于计算机的培训教程。

由于这些都是新技术,因此有结构的、有督导的练习是访谈员训练中最重要的组成部分之一。让访谈员轮流扮演受访者和访谈员的角色,这是很常用的方法。练习应包括争取合作及处理问答的过程。对那些由访谈员不认识的真实的受访者参与的访谈实践进行监控,这也是很有价值的。对于个人访谈,督导者可以陪同或观察新访谈员练习访谈,或者检查访谈的磁带录音。在电话访谈中,可以直接对访谈进行监控,或者先将访谈内容录音,以后再进行检查。

有两个研究(Billiet & Loosveldt,1988;Fowler & Mangione,1990)表明,如果培训时间不足一天,则不能培训出令人满意的访谈员,访谈员不能按指示完成任务,并对调查资料会有不利的影响。专业调查机构的培训课程一般持续 2 ~5 天。培训时间长短取决于多个因素,其中包括要培训的访谈员的总人数和培训内容的复杂程度。但是,对培训质量最关键的很可能是在督导下进行的实际访谈的数量。

监　督

有效监督的关键是获得评价访谈员表现所需的信息,及投入评价信息和提供及时反馈要求的时间和资源。对访谈员的表现进行督导主要包括四个方面:成本、回答率、已完成问卷的质量以及访谈质量。较之其他访谈方式,电话访谈的督导比较容易,因为它可使用中央控制设备。

成本监督

对访谈员的成本进行监控,需要及时了解访谈花费的时间、工作效率(通常是完成的访谈)以及每英里访谈用车的费用。高成本的电话访谈员可能工作效率低下、被拒绝率高(一次拒绝所花时间与一次访谈所花时间几乎差不多),或者找理由(如编辑访谈资料、削铅笔等)每小时只打几个电话。高成本

的入户访谈的访谈员的居住地可能与样本所在地较远，这样就可能因为路途太远，而使访谈时间太短，或在不恰当的时间进行访谈（周末和晚上显然最有效率），或只有很低的回答率。

回答率监督

对访谈员的回答率（特别是拒绝率）进行及时监控是十分重要的，但这并非是一件易事。主要有以下三个问题：

- 对于个人访谈而不是计算机中央控制的电话访谈来说，要及时掌握访谈员的结果信息是较难的。
- 访谈员可把失败的访谈归入其他类别，低报拒访的人数。
- 给个别访谈的访谈员分配的访谈任务没有可比性，因此每个访谈员的拒访率并非是反映调查员之间访谈实施情况差别的一致性的指标。这一问题在中央控制的电话访谈中就不存在。

调查结束前是无法准确计算出回答率的。但是，访谈员在数据收集过程中努力识别拒绝者可以提醒督导员注意这些问题，因而也是访谈监督的一个重要部分。很难帮助那些有回答率困难的访谈员。在电话调查中监督员可以听介绍，并在访谈结束后（或在访谈中）立即就访谈员如何提高效率提供反馈意见。对于面对面的访谈来说，监督工作更困难，因为除非督导员与访谈员一起去，否则他无法观察到访谈员所运用的访谈方法。因此，督导员通常只能通过听访谈员介绍来对访谈员的情况作出判断。

督导员可以给访谈员提供有益的建议。确保访谈员能对调查有比较全面的了解，这十分重要。让访谈员练习简明扼要地回答调查中常常遇到的问题是十分有益的。除了将介绍工作做细致外，督导员可能还需要处理访谈员对于与人们联系或对于研究项目及其价值的一般感受。但是，再培训的作用也是很有限的，总是会有那么一些人，无论怎么进行再培训，也无法使他们的回答率达到令人满意的程度。而保持高回答率的最

有效的办法是将这些低效的访谈员从研究中除名,这种做法令人不快,但我们别无选择。

复查已经完成的调查工具

如果访谈员使用纸笔调查工具,就应当对完成了的调查问卷样本进行复查,以便评估访谈员收集到的资料的质量。在复查一个已完成的调查时,显然能够查看记录是否清晰、跳答指令的运用是否得当、所得答案是否完整到可以进行编码等。另外,查看一个已完成的访谈,通过比较记录摘要或阐释,能很好地看到访谈员在多大程度上逐字记录了回答者的答案。对计算机辅助访谈而言,这些问题(除了与陈述答案相关联的记录和探查之外)是无关紧要的。

问答过程监控

通过复查已完成的调查工具无法监控访谈质量,复查已完成的调查工具并不能告诉督导员有关访谈员开展访谈工作的方式,也不能告诉督导员那些答案是如何得来的。督到员只有直接观察访谈过程,才可能了解到这些。

中央控制的电话调查允许直接监控访谈员收集数据的方式,督导员能够也应该随时监控访谈员。督导员应该系统地听取每个访谈员对一个样本进行的访谈的全部或一部分,从而评价(还包括其他内容)研究介绍语的恰当性、按照书面文字提问的严格性、探查的适当性和间接性、处理访谈中的人际关系的恰当性等。如果能在屏幕上对访谈员诸如此类的各方面工作进行例行的评价,那么问答过程便臻于完美(Cannell & Okesenberg,1988)。

如果访谈员是在受访者家里或其他较远的地方进行调查,问答过程就难以监督。只有两种解决办法:督导员以观察员身份陪同访谈员,或者对访谈进行录音。如果不录音或观察,研究者就无法评估访谈质量。所有测量过程中最重要的方面都没有监控,因而不能辨别出不合格的访谈员以便对其进行再培

训,研究者除了说已经告诉访谈员该做什么之外无法再就访谈的质量报告任何东西。实际上,从访谈员的角度看,如果标准化访谈只是作为培训的重点却没有进一步深入实际,就很难让人相信它的重要性。福勒和曼吉温(Fowler and Mangione,1990)提供的证据表明,如果不用磁带录音直接监控个人访谈员,他们就不太可能以培训所学的方式进行访谈。福勒和曼吉温(Fowler and Mangione,1990)以及比里特和洛夫特(Billet and Loosveldt,1998)都发现,如果以这种方式直接监控访谈员,资料质量就能提高。现在很清楚,对访谈程序进行直接监控应当是管理得法的调查的一个组成部分。

调查问题

尽管培训和督导对令人满意的访谈来说很重要,但研究者能采取的最好措施是设计出良好的调查工具。研究表明,某些问题总是被误读,有些问题的回答总是不完整而需要访谈员去探查以得到完整的答案(Fowler,1991;Fowler & Cannell,1996;Fowler & Mangione,1990;Oksenberg et al.,1991)。运用第6章阐述过的前测法可以鉴别这些问题。

访谈员要探查、解释或澄清的越多,他们对答案的影响也就越大。调查工具越完善,访谈员就越有可能开展标准化的、令人满意的访谈。福勒和曼吉温(Fowler and Mangione,1990)及福勒(Fowler,1991)详细讨论了良好的问题设计在令人满意的访谈中的作用。

访谈过程

培训和激励回答者

研究证明,良好的问题设计在促进访谈标准化方面有重要作用(Cannell et al. ,1987;Cannell,Oksenberg & Converse,1977;Miller & Cannell,1977)。例如,研究者可以帮助访谈员用一致的方式培训受访者。访谈开始前,访谈员可以读出下列内容:

> 由于大多数人还没有接受过这种调查,所以在开始调查前,我需要先给你说一些有关访谈过程的事情。调查中将问你两种问题,一种问题需要请你用自己的话来回答,这时我会记下你所说的每一个字,不做任何概括;另一种问题,我们会提供一些答案,请你从中选出与你的想法最接近的选项,尽管其中的任何答案都不可能与你的想法完全一样,但还是请你选出一个最接近你想法的选项,以便我们可以比较你与其他人的答案。

有趣的是,访谈员很喜欢用这样的介绍,因为它介绍了受访者要做的事,使问答过程更顺利。实际上,优秀的访谈员能够自己做出这样的介绍。提供清楚的介绍,让所有访谈员做同样的事,这样可以减少差异。另外,这种介绍对访谈员的成绩也有积极影响。一旦访谈员读出这些介绍,清楚解释了访谈要达到的目标,访谈工作就易于按照要求的方式进行了,并且不易出错,因为此时受访者也知道访谈员希望他们做的是什么了(Fowler & Mangione,1990)。

给受访者的标准化介绍也可用做为操作确定目标和标准:

> 你尽可能准确地回答问题是非常重要的,不必匆忙。如果必要时你可以查阅有关记录。如果有什么问题不清楚,可以要求我进一步解释。

诸如这样的陈述可以保证受访者对自己优先考虑的事情有共同理解。有的访谈员无意识地向受访者允诺,只要他们参加访谈,自己就会使访谈变得容易。这些心急的访谈员向接受访谈的人传递了一种信息,那就是速度比准确度重要。如果我们向回答者宣读的是如上面讨论的那种引导语,它就会促使访谈员和受访者把答案的准确性和数据的质量视为访谈中最为重要的问题。这样访谈员之间的变异源减少了,而访谈员与受访者二者按规定完成访谈的可能性却提高了。

坎内尔、沃克森伯格和康弗斯(Cannell, Oksenberg, and Converse,1977)提出了一种更好的方法。在允许受访者参与访谈前,要求他们签字承诺尽力给出准确而完整的回答。本来预期会有很多人拒绝这样做,但实际情况却并非如此。尽管回答率没有受这种形式影响,但是报告的质量却提高了。在电话访谈中用口头策略进行同样的操作,也证明能有效提高报告的质量(Cannell et al. ,1987)。

坎内尔还试图将访谈员要给受访者的强调的内容的用语标准化。访谈员经常会在无意中强调了一些并非期望的受访者行为(如快而不假思索的答案。Marquis, Cannell, & Laurent, 1972)。坎内尔、沃克森伯格和康弗斯(Cannell, Oksenberg, & Convers,1977)指出,如果预先设计的访谈计划能促使访谈员对受访者的良好行为(如检查以往的记录或慢慢地进行答问)进行鼓励,那么受访者的报告质量就会提高。平常运用这些程序有些难度,但这项工作强调的是把访谈员做的不恰当的提示降到最低程度。

总之,除了直接的问答过程之外,还有许多重要的工作需要访谈员完成。特别是访谈员要负责告诉受访者访谈的进程:受访者应该做什么、访谈员要做什么、双方的共同目的是什么。这个方面的工作主要是由访谈员自己来做,而不同的访谈员做这些事情的方式不同,这必然会给数据带来影响。通过为受访者开发标准化的介绍程序,研究者可以使访谈员的工作更容易,减少访谈员之间的变差源,改进访谈双方的行为,使测量过

程变得更加完善。

标准化措辞

前面提到过，严格按照写出的问题提问是标准化测量的基础，但并非每个人都同意这个观点（Tanur，1991）。标准化访谈的批评者注意到，并非所有受访者对某些问题的理解都是一致的。他们认为，在这种情况下，让访谈员自由澄清或解释问题将提高数据质量（例如，Schober & Conrad，1997）。同时批评者指出，如果访谈员只用预先写好文字，有些数据收集任务（如，从几个不同的人或者事中收集相同的数据）会使调查员和被调查人之间的交流变得十分生硬或别扭。批评者认为，在这一事例中，如果给予访谈员在话语上更大的灵活性，就可以使访谈双方的交流变得更加愉快（Schaeffer，1992）。

某些对标准化访谈的批评，主要是集中在设计低劣的问题造成的结果上（参见 Suchman & Jordan，1990）。在问题表述不清，或者提供给访谈员的脚本很难使用时，最好的解决方法是重新写出更好的问题，而不是让访谈员重新设计问题（Beatty，1995）。确有证据表明，如果访谈员可以灵活自如地重述或解释问题，他们常会改变问题的原意，从而使数据的质量更糟而不是更好（Fowler & Mangione，1990）。然而，确实有这样一些问题——诸如一系列重复的问题，或那些本来只需给少数受访者详细解释的而对所有的受访者都进行解释则会令人讨厌的问题——若能给访谈员更大的灵活性，则可以得到更好的解决。不仅如此，即使在访谈员改变问题的措辞时，得到的证据并未一致证明与访谈员有关的误差或回答误差一定会因此而有所增加（Dykema，Lepkowski，& Blixt，1997；Fowler & Mangione，1990）。

人们一直在进行给访谈员更多提问和探查答案的处置权的实验（Conrad & Schober，2000；Schober & Conrad，1997）。迄今为止，实验结果也见仁见智，众说纷纭：有些报告的准确性可能有所提高，但访谈员的培训量及访谈时间却明显增加了。在

什么时间、以什么方式给访谈员更多的灵活性还有待进一步实验。而无论什么时候对于大多数调查来说，设计出访谈员能够且愿意严格按所写形式提问的问题，仍然是做好一个调查的主要方法。

检验访谈的真实性

访谈员有可能在访谈中作假，这个问题应该引起我们的关注。这种事情发生的可能性，随样本、访谈人员的具体情况和实际访谈过程的不同而不同。这样的情况大多数发生在入户访谈，或访谈员在自己家中进行电话调查时。在这种情况下，督导员无法观察到数据收集的实际过程。实际访谈需要花费很多的时间这使得有些访谈员不去花费很多的时间和精力做访谈，而是去造假。

从长远来看，预防访谈做假的最好办法，也许是建立一支对研究质量和研究机构负责的访谈员队伍。新雇佣的访谈员最有可能做假，但是，即使是拥有经验丰富的专业化工作人员的机构，也需经常对访谈样本做定期检查，以确保访谈得以真正实施。

进行真实性检验的方法有两种：一种是在事后，给所有的受访者寄一封简短的追踪问卷，了解他们对访谈的反应。另一种，也是更常用的办法，要求访谈员获取每个受访者的电话，然后督导从已经完成的调查中，抽取一个样本打电话进行检查。只要事先让访谈员知道，在调查完成后将会通过邮件或电话对访谈的真实性进行检查，便会在一定程度上阻止访谈员在访谈中的造假行为。此外，我们还可以说，只有能够说访谈资料是经过核实了的，进行诸如这样的检验才可以使数据的用户不再对数据的真实性心存疑虑。

访谈过程对调查误差的影响

本章开始就提到,访谈员会对回答率、报告的准确度以及测量的一致性或精确度有所影响。而这些影响中的每一种都会在调查估计值中有很重要的作用。

好的调查管理最易观察到的效应之一是回答率。尽管第3章已详细讨论过这个问题,但仍需重申的是,访谈人员的素质对任何调查所能得到的回答率都是至关重要的。

测量访谈员在问答过程中造成的误差是比较困难的。在很多时候,这样的调查误差是很难觉察的。在对主观状态提问时,对偏倚或不准确性进行客观检查一般都是毫无意义的,这一点在第5章已经讨论过。但在有些研究中,研究者对要求受访人报告的事实使用了客观的测量,这样的测量允许对报告的精确度进行评估。在一个诸如这样的研究中(Cannell, Marquis, & Laurent, 1977),对那些一年前家中有人住院的家庭的样本进行了访谈。这样的调查报告的精确度,可以通过对健康访谈的住院情况报告和医院的有关记录加以比较来评估。报告的精确性的一个量度其实就是调查报告的已知的住院百分比。

在这个研究中,我们发现分配给每个访谈员的访谈数与在访谈中未报告的住院百分比高度相关($r = 0.72$)。分配的访谈人数多的访谈员,不管出于什么样的原因,他们收集到的准确数据都比分配的访谈人数少的访谈员少很多。

另一个研究运用同样的标准(报告的住院百分比;Cannell & Fowler, 1964),也得出了相似的结论。在本研究中,让访谈员的受访者中一半一次性地报告住院情况,而让另一半在访谈员结束其他健康访谈离开后自我填答关于住院情况的表格。研究发现,受访者报告的住院准确率高的访谈员,其受访者在访谈员走后做自我填答式问卷时也做得令人满意($r = 0.65$)。这一研究结果表明,访谈员不仅对受访者的报告误差有很大影

响,而且会在对回答者的激励方面对回答者能否很好地回答问题有很大影响。在这两种情况中,访谈员对报告准确度的影响都是明显的。

不对数据进行检验,我们就无法对准确性作出评估。然而,对访谈员对受访者的答案的影响程度作出评估则还是可能的。如果所有访谈员都以非常标准化的方式做访谈,那么就无法用不同的访谈员来解释答案中的任何变差。如果可以用不同的访谈员进行的访谈来部分地预测答案的话,那么我们就有理由认为访谈员可能在一定程度上不恰当地影响了答案。格罗夫斯(Groves,1989)对访谈员对问题的答案的影响程度的计算方法进行了详尽的讨论,并对许多来自访谈员对研究结果的影响的研究技术按结果做了概括。研究结果表明,对于访谈员所问的很多问题,并不能看出访谈员对答案有任何影响。但在大多数调查中,有大约 1/3 到 1/2 的问题,其答案受到了访谈员的严重影响。

这些访谈员的影响导致了调查估计值的标准误差的增加。增加的大小取决于类内相关系数(rho)的大小和访谈员分配到的任务数量的平均值的大小(参看 Groves,1989;Kish,1962)。如果类内相关系数是 0.01(Groves 发现这基本上等于平均值),每个访谈员平均做 31 次访谈,均值的标准误差就会比样本设计的估计值高 14%。当访谈员的平均访谈数接近 50,那些类内相关数是 0.02 的题项的标准误差的估计值会增加 41%。

从这些讨论中,可以得出以下几个有关访谈员在调查数据的总误差结构中的作用的结论:

- 访谈员不仅对回答率会有影响,而且还会对受访者在调查过程中给出的不准确答案的程度和测量的不一致性有影响。现有证据明确显示,访谈员是多种测量的一个重要的误差源。
- 访谈员接受的培训和督导能极大提高访谈员的一致性,从

而提高估计值的可靠性并降低偏倚。而那些只受过少量培训（例如，不到一天）或只得到过少量有关访谈质量反馈的访谈员，一般都是比较差的访谈员。

- 那些用于建构受访者培训和对回答者进行引导的程序，能把访谈员不合理的反馈降到最低。而一般地讲，更多地控制访谈员的行为能减少访谈员对调查数据的影响，进而提高总的准确度。
- 良好的问题设计是做好访谈的关键。
- 一个未被重视的设计选项是访谈员的平均任务量。尽管访谈员用得少，其培训和管理成本也会降低，但如果一个研究者让一个访谈员承担过多的访谈任务，那么他就将要为此而付出为数据的可靠性下降的代价。减少访谈员的平均工作量，从而提高调查估计值的精确度，这是最有成本效益的方法。
- 实际上，几乎所有关于调查估计值可靠性的报告都忽视了访谈员对数据的影响。这部分是由于以便利性和地理邻近性等非随机方式给访谈员布置任务时，研究人员无法从抽样影响中分离出访谈员的影响。但是，对于大多数调查项目来说，访谈员的影响是误差的重要来源，任何忽视访谈员影响的关于调查估计值精确性的报告，都会低估调查误差。

总之，访谈员在调查数据误差中的作用未被普遍重视。尽管绝大多数调查研究者都知道访谈员需要培训，但是培训和督导访谈员的程序相差很大，而且不够充足。除培训和督导之外，在尽可能地降低访谈员影响方面，研究者往往没有再做其他任何努力。但是，调查设计的这些方面是提高调查数据质量的最有成本效益的方法，因此我们应该在调查设计和研究中充分重视访谈员对调查估计值的影响。

练　习

用录音带录下你或其他人使用标准访谈计划时所进行的角色模仿访谈(第6章中提出的问题,或其他渠道的计划)。然后听录音,通过记下每个问题所出现的下列错误来系统地评价访谈员的表现:提问没有完全按书面文字进行;探查不完整答案时采用了有偏的(直接的)方式;没有探查不明确的答案;其他任何可能带偏倚或不标准的人际交往行为。如果是以集体方式进行评价,就特别有指导意义,这样可以对访谈员的误差进行讨论。

第8章　整理调查数据

调查答案通常都会被转换成用于计算机分析的数据文件。本章将介绍数据格式的选择和使用、编码设计、编码程序和管理方法、数据输入和数据检验程序。

在完成调查数据收集之后,不论采用的收集方法是什么,它们都要被转换成宜于用计算机进行分析的格式。本章主要讨论如何对完成的问卷和调查访谈进行适当处理,并转变为计算机能阅读和处理的格式的过程。编码或数据整理的过程牵涉到以下五个独立的阶段:

- 确定格式(将数据组织成文件的方式)。
- 设计编码(对回答者的答案赋值,以便能用计算机处理)。
- 编码(将答案转换成标准类别的过程)。
- 数据输入(将数据变成机算计机可阅读的格式)。
- 数据清理(在开始数据分析前,对数据文件的准确性、完整性和一致性进行最后的检查)。

在将答案输入数据文件时可能会发生两种误差:第一种是抄写误差,任何人在抄写过程中的任何时候都有可能出错。第二种是编码错误,在编码时给答案选择了不恰当的码值。数据录入质量控制方法的选择,因选择的特定的数据输入和编码程序而异。接下来就将讨论这些问题。

编排数据文件的格式

每种分析软件对数据文件的格式都有自己的要求。在设计数据输入过程之前,最重要的一步是确定使用的分析程序,及它们对文件格式和丢失数据处理的特殊要求。“记录(record)”一词在这里特指所有的数据都属于某一个体、个案或访谈。一个记录由单独的一行或若干行组成。在历史上,一张80列的卡片作为一个记录数据的单位,数据记录的方法是在卡面上打孔。一张这样的卡片相当于一行数据。现在虽然一般都把数据放在硬盘或软盘上,但通常每个记录或访谈的数据仍然以一组有序的行的形式储存。尽管协议和规则会因使用设备和程序的不同而有所不同,但是有些问题却具有普遍性:

- 虽然实际上已经不再使用以往那样的数据卡，但是用卡行的方式来确定数据的具体位置的方法却沿用至今。在这种格式中，80 列为一行，每满 80 列就必须新起一行。而现在很多分析程序已经可以很轻松地处理传统的行长超过 80 列的数据。如果我们使用的是这样的程序，那么就没有必要受卡行格式的限制。
- 回答者的系列标识符都位于每一记录或访谈的每一行中的同一位置。令行或卡号位于数据文件中的每一行的同一位置，有助于检查数据文件的完整性。如果对这些标识符做了排序，它们便可使数据保持一种次序，而数据的次序则对数据文件的完整性的检查至关重要。
- 如果按照在调查工具中出现的顺序对其数据进行编码，那么编码、数据输入和编程工作变得比较容易。这样不仅能减少误差，而且也为我们提供了一种免费的控制数据质量的手段。
- 有些程序，在一个字段或列中也可以接受多个编码，但有些程序则不能。但是最好的方法还是在一个容纳数据的字段单独放一个正的数项。同样，有些计算机程序将空白理解为零，而另外一些程序则不这样解释。如果零是有意义的，那么最好是将它的编码定为零，而不是留下空白。如果想要为无回答编码，就应该使用某些特定的值（通常是数字形式而不是空白）。

构建编码

编码就是一套把答案转化为数字或者把数字转化为字母的规则（一些系统接受字母形式的值，但是调查中的绝大多数编码都只使用数字码）。哪个答案用哪个数码并不取决于计算机。而编码不能模糊不清这一点，却是可靠地进行编码和恰当

地解释数据的关键所在。应该用一套明确的规则来确定应把什么样的数字分配给每个和所有的答案(或其他形式的结果)。此外,我们还可以通过编码设计,最大限度地减少发生在编码和分析过程中的误差。以下是一些常见的编码设计中应遵循的原则:

1. 确定用于那些没有答案的问题的各种丢失码。这些编码要能区别以下不同情况:

- 模糊不清的信息(not ascertained information),指那些由于访谈员或者回答者的原因而导致的本来可以得到的可编码的而最终未能得到的信息。某些研究者喜欢用不同的编码来把它们与回答者拒绝回答的问题和因其他原因没有回答的问题加以区别。
- 不适用的信息(inapplicable information),指对某些特定人群不适用的信息(例如:向没有住院的人询问住院时间)。
- "不知道(don't know)",这类答案也许也可当作一种模糊不清的信息,或特殊的类别的数据丢失。

2. 分配的数码(number)必须首尾一致,对"模糊不清","不知道"或"其他"这样的答案必须使用相同的编码。编码越一致,编码员和程序设计员产生的误差就越少。

3. 尽可能让编码与现实中的数字一致,编码的数值要准确(例如:把45周岁编码为45)。同时,如果没有特别的原因,最好以答案在调查中出现的顺序对一份答案进行编号。

如果提供给回答者的答案选项或答案形式是高度结构化的,那么,编码员的工作就只是为给定的一系列答案指定数码和说明缺失的数据。但是,如果回答者被要求用自己的语言回答问题,那么我们就无法预先确定答案的范围。对这种开放式回答的问题,编码的开发就是研究者对得到的答案的类别进行鉴别和排序的过程。

所谓编码开发就是创建一些能把在分析上具有相似性的答案归类的类别,进而据此区别不同类型的答案。如果分类太细,其结果就是许多类别里只有很少的条目,因而难以进行分析,从而白白浪费了编码员的时间;另一方面,大的、过于笼统的分类则会掩盖答案间重要的差异。衡量编码好坏的一个标准就是给每个答案分配一个,且只分配一个唯一的编码数值。另外一个标准就是把答案归入有分析意义的类别中。但只有在有了明确的分析计划时,我们才能对后一个标准的好坏作出评估。

为了构建这样的编码,我们必须:

- 了解答案的哪些特点具有分析意义。正确的做法是,先记下那些在研究者看来是很重要的、每个问题的答案间的各种差异。
- 把早期访谈得到的答案制成表格。然后编制一个编码草稿,以便给这些答案分类。
- 把这个分类方案用于 10 ~ 20 个其他的访谈中,必要时则予以修改。
- 给那些不能清楚地确定其类别的答案以单独的一类编码,即"其他"。让编码员对这些答案做出注释。这些注释在可以用来扩充和澄清编码,或在增加必要的类型的同时,也提供了一种包括在"其他"类中的答案记录。
- 为了使编码员能就编码中存在的问题和模糊不清的定义与研究者进行交流,也可以使用同样的注释。这样研究者可以根据这些注释对有关的定义和规则做必要的改进。

这些步骤加之对编码进行有效的检查(下面将提到),会产生穷尽而互斥的类别系统。这一系统可以明确地将每个答案放进一个,且只放进一个位置,而这一位置则为编码员、编码督导,以及将要进行数据分析的研究者所共有。

编码和数据输入方法

多年来,从自动制表开始到20世纪70年代末或80年代初这段时期里,从问题回答到资料输入的过程,一般包括三个步骤:将答案记录在纸质的调查工具上;经过培训的编码员将答案转变为编码数值,并将数值写在专用的编码单上;和打孔操作员将编码的数值打到80列的IBM卡上。

即使现在我们已不再按照这些步骤进行实际操作,但是理解那些设计用来最大限度地减少由这些步骤引起的误差的质量控制步骤的目的,仍然是很有意义的。

访谈员记录:没有切实有效的方法来检查访谈员是否准确地记录了答案(例如:是否勾了正确的方框)。然而,尽可能地将需要由访谈员确定其编码的问题数减少到最小的程度,则不失为一种行之有效的降低这类误差的好方法。对于开放型的答案最好先用文字逐字记录,然后再由训练有素的编码员在督导员的指导下对它们进行编码。在遇到根本不清楚怎么确定分类的时候,应该要求访谈员随时记录下所有相关的资料,这样就可以在编码时对它们一一进行斟酌,并在通盘考虑之后作出一致的处理。

编码:编码过程的质量控制包括以下几个方面:

- 培训编码员,包括让所有的编码员对同一个调查回收的若干问卷进行编码,然后比较编码的结果,以确定他们是否使用了同样的编码方法。
- 对每个编码员的工作进行单独的抽样编码检查。这样做可以达到两个目的:①确定哪些编码员编错了码;②找出编码规则中存在的模糊不清之处,在必要时对它们作进一步的澄清。
- 应该做出一个规定,要求编码员在遇到不知道怎么进行编

码的答案时，如实记录。督导员应对这些记录进行例行的评阅。这些记录是检查系统的延伸，它有助于督导员确定哪些编码员和哪些编码需要他们予以关注。

数据录入：新的电脑技术给数据录入过程带来了巨大变化。前面提到的旧技术使用的是打孔机，它通过在 IBM 卡上打孔记录数据。这种机器功能单一、价格也比较昂贵，因此只有专用的打孔设备才可能使用这样的技术。由于电脑相对廉价和功能多样，而且它无处不在，并且任何电脑都可以用来录入数据。因此，不用再将数据送到专门的打孔场所，就可以由任何人在任何地方录入数据。此外，电脑可以用编程来改进数据录入过程：

- 任一字段都只允许录入合法的码值。
- 检查当前字段录入的码值，确认它们与以前录入的数据是否一致。
- 能自动确保对列联问题(contingenly questions)进行恰当的处理(即在一系列问题只问及部分回答者时，能对列联指令进行编程，从而令计算机自动地在那些跳过的问题的字段中填上适当的编码)。

尽管这些检查不能识别那些没有违反那些已经编好程的规则的数据的输入误差，但是它们能使很多的数据输入误差，可以在它们立即就能得到改正的时候被发现。

伴随这种新技术的出现，目前产生了两种比旧的三步输入法更为常用的资料输入法。第一种是两步法，也常常称之为直接数据输入法：这种方法的特点是在访谈员或回答者填写书面形式的调查工具之后，立即进行编码和数据输入工作。即填写、编码和录入是在同一步进行的。然后再让第二个人再一次独立地对数据进行编码和输入，借此对前一步的编码和录入的数据进行检查。

另一种叫做二步光学扫描法的编码和数据输入法也值得一提。现今扫描技术的发展可谓日新月异，扫描技术用于数据

输入有两个方式：

一是对备择答案选项用条形编码进行编码，这样我们就可以用扫描仪扫描所选答案旁的条形码来输入数值。这种方法的优点在于，即使是那些没有数据录入技能的人也可以胜任数据录入工作。

二是扫描诸如标准化考试之类的特制的答题纸或者表格。这种已经存在了好多年的做法可使数据输入的成本降得很低。其成本基本上等于设备购置费，如果需要使用特定格式，那么除设备费之外，还要加上设计和印制的费用。

从历史发展来看，光学扫描仪在调查中的使用率呈下降趋势，这是因为：

- 格式不便于用户使用，而调查研究者则希望问卷尽可能地便于使用。
- 为比较小的调查建立特定的表格成本比较高。
- 可能导致严重的数据丢失，在缺乏积极性或未受过训练的回答者使用这种方法时，情况尤其如此。

用手工检查缺失题项以识别机器读不出的符号，这种方法可以解决最后一个问题。但是，要使光学扫描仪得以在调查中得到更为广泛的使用，还有赖于技术的提高。与以前的扫描仪相比，现代化的扫描仪更能兼容不完整的符号。与此同时它们也已经是可以用于一系列的格式的调查工具，从而使自己变成一种易于用户使用的调查工具。

尽管技术进步到了可对手写的答案先进行扫描，然后再由编码员对它进行编码的地步，但扫描仪只有在对固定选项的预编码数据才能进行正常的扫描。而比较好的设备现在仍然比较贵，因此扫描仪在数据输入中的作用只有在将来才可能会逐渐增大。迪尔曼和米勒（Dillman & Miller，1998）、迪尔曼（Dillman，2000）以及布洛姆和赖伯格（Blom & Lyberg，1998）很好地总结了现行扫描方法及其局限性。

比旧三步法更常用的第二种方法是一步法，它又被称为电

脑辅助电话访谈法（CATI）、电脑辅助个人访谈法（CAPI）或者电脑辅助自填问卷法（CASI）。它要求访谈员或者回答者直接把答案输入电脑，然后在需要的地方进行编码。访谈中不使用纸和笔。从数据输入的角度看，一步法与在英特网上收集数据的方法基本相同。在这两种方法中，回答问题和数据输入是同步进行的。

正如在第 4 章所述，电脑辅助资料收集方法越来越普及。对于电话访谈来说，每个访谈员需在电话局有个终端。终端屏幕上显示出问题，访谈员要读出问题让回答者回答，然后访谈员将与答案相对应的数值输入终端，而数值的输入会引发下一个问题出现在屏幕上。研究员可以给电脑编程，让它只接受合法的输入并检查输入的答案是否与以前输入的数据一致，这样访谈员能够清楚回答者答案里明显不一致的地方。手提电脑或其他可以携带的个人计算机也为家庭访谈提供了同样的数据收集方式，这与用网站进行数据收集的方式是一样的。

所有使用电脑辅助的数据收集系统，都有以下几个优点：

- 电脑可以处理在纸笔调查中，访谈员或回答者则难以处理的复杂的问题模式。
- 在问题措辞或一系列提问中，可以把从先前的问题甚至先前的个案中获得的数据考虑进来。
- 如果给出的数据不一致，访谈员可以立即纠正。
- 电脑能将数据添加到立即就可以进行分析的数据文件中。

也有不使用这个系统的恰当理由。其中一个可能是最重要的理由是必须花时间给电脑辅助方案编程，而且程序必须准确无误。因为在进行访谈时，访谈员无法像处理印刷错误一样处理程序错误。在没有访谈员时，回答者直接输入他们的答案就有可能出现更多的错误。尽管那一些简单的只有比较少的跳答题的调查工具，一般才需要进行太多的程序测试，但是在访谈正式开始前，仍需用相当多的时间来进行程序测试和修正。

另外,整个数据录入过程中缺乏质量控制。除了确保输入的是合法编码并具有内在一致性之外,访谈员使用电脑辅助系统进行的任何数据输入或任何编码决策都没有经过检查。尽管出现打字错误的几率相对来说比较低,但编码决策的质量却需要特别注意(Dielman & Couper,1995)。由于对编码决策缺少控制,当提开放式问题时,CATI 和 CAPI 的访谈者通常都要将答案逐字输入电脑,以备后来进行编码。尼科尔斯(Nichols,1988)、贝克和莱弗(Baker & Lefes,1988)、萨瑞斯(Saris,1991)、卡特林和英格拉姆(Catlin & Ingram ,1988)、尼科尔斯、贝克和马丁(Nichols,Baker,& Martin,1997),特别是还有库珀等(Couper et al. ,1998)都曾编辑出版过有关的专著,对电脑辅助系统的特点、使用方法和经验做了详细的总结。

数据清理

一旦完成对访谈的编码并把数据输入磁带或者磁盘文件存储后,此时就需要对数据进行检查。最重要的检查是弄清数据文件是否完整和有序。此外还要对每个字段进行检查,以确保其中没有非法的码值。即使在数据输入时已经进行过检查,仍有必要看一看数据的总体分布,以确认一切是否都是按计划进行的。当然,如果在数据输入时没有进行过检查,那么除了这些检查之外,我们还需要对数据的内部一致性进行检查。

发现错误时就要诊断其来源并进行更正(注意:在使用 CATI、CAPI 或者 CASI 系统时不可能做到这一点,因为没有保留可以确证的答案脚本)。由于在改正过程中有可能会再次出现错误,所以在更正之后还应再检查一遍。如果数据量很大,这个清理过程会很耗时并且容易产生误差。最为理想的是能够在数据输入时就发现错误,这样就可减少对后续数据清理的依赖。

编码和资料归并导致的调查误差

由于编码和数据归并能在严密的督导下进行,并能得到全面彻底的检查,这就使它有可能成为调查过程中几乎没有误差的一部分。不仅如此,编码和数据归并的成本通常只占整个调查成本的一小部分。

在处理封闭式答案时,数据输入过程中误差率不到1%。如果那些数字是直接输入并全部经过核实,这就等于对抄写过程本身也做了检查,最后得到的数据的误差率会更低。

开放式答案编码的信度会因问题的质量、编码的质量和对编码员的培训和督导不同而有所不同。如果研究者有好的问题,并且编码类型都很清楚,那么就可以预计编码有90%以上的信度,也就是说编码员和编码检查员之间出现分类不一致的比例低于10%。没有受过训练的编码员和不对编码进行恰当的检查可能会使误差率明显上升。编码有赖于对完整的定义的了解,例如,职业类型、健康状况或特定的犯罪等,因此研究者也许对编码员的培训和对编码的检查予以特别的关注。

经过仔细检查,数据输入过程几乎可以达到没有错误。虽然有一些操作员能以相当高的精确度(出错率低于1‰)输入数据,但是仍然不能想当然地认为,数据录入的准确度一定会达到那样的水平。

对编码和数据输入过程的选择,通常出于其他的考虑,而不是对使误差最小化的考虑。在访谈中构建文档的速度和抓住错误的机会,是CATI和CAPI系统吸引人的地方,这就像电脑辅助使问题措辞具体化和使提问有序化的长处一样。但是,单从减少误差这个角度来看,在调查需要做大量的编码工作时,那种由编码员直接输入数据,并对录入的数据100%进行校验的二步输入法可能是最佳的选择。除此之外,我们找不到任何其他的系统,能够对所有的编码和所有的数据输入进行真正独立的检查。

第9章　调查研究中的伦理问题

跟所有的社会研究一样，调查研究应该力求以一种避免给参与者、回答者和访谈员造成危险的方式进行。本章将简要阐述怎样使调查合乎道德规范。

在以人为研究对象的所有调查中，研究者都必须注意研究过程中的道德规范问题。其中最为重要的是，研究者必须确保调查结果不会对任何个人造成任何不利的影响。不仅如此，一个优秀的研究者还应尽可能地扩大研究过程中的积极成果，而这在某种程度上讲是完全可以做到的。

在美国，几乎所有的大学和其他很多机构，在进行联邦政府赞助的调查时都有一个体制审查委员会（IRB），负责监督以人为对象的研究。在建议进行一项研究时，调查主持人在开始收集资料前，必须把议案交给 IRB 进行审查。

IRB 审查的目的是为了保护研究对象、研究者和研究机构。他们最关心的问题是研究是否会使参与者冒一定的风险。只是涉及人类的研究活动……，教育考试、各种调查程序或者对公众行为的观察可“免于”（审查），除非：

- 以可以识别研究对象的身份的方式记录信息……
- 公开研究对象所提供的任何答案……可能在一定程度上使其陷于不道德的境地，或违背公民责任，或有损其经济地位、就业或声誉（健康和人类服务部，1996，P. 5）。

根据这些审查方针，即使不是大多数调查，至少很多调查在技术上可免于（审查）。然而因为调查毕竟要涉及被调查者，调查的负责人还是须向 IRB 提供有关材料，以便他们（通常是委员会主席）确定调查方案是否符合有关标准。如果调查确实存在某种潜在的风险，或者牵涉到易受伤害的群体，IRB 有责任对所有程序进行正式的审查，以确保研究对象能得到很好的保护。

本书不可能阐述所有涉及特殊群体研究中的问题。当对儿童、智力发展迟缓的人、精神病患者、罪犯或其他特殊总体进行研究时，研究者应从其他地方获得指导。希伯（Sieber，1992）更详细地论述了怎样以道德的方式收集资料。下面是一些对一般总体进行调查时应遵循的道德原则，这是所有调查研究者都应该熟悉的。

告知回答者

一般说来,调查研究过程需要调查对象的自愿合作。合乎道德的调查研究的一个基本前提是应该告知回答者他们自愿参加的是什么活动。具体来说,回答者在回答问题之前应该知道以下信息:

- 进行这项研究的机构名称。如果有访谈员参与,回答者也应该知道访谈员的名字。
- 主办人,即那些支持和资助研究的人。
- 对研究目的作准确而又简明的阐述。研究是否试图增进一般的或基本的知识?或者是否有些研究援助的计划或行动过程?这个研究覆盖的问题或主题是哪些?研究中设计了哪些要回答的问题?
- 对答案的保密程度作精确的陈述。如果调查提供的保密性有某些限制,研究者一定要清楚地告知回答者。
- 保证合作完全出于自愿,即使不参与调查也不会由此而产生任何负面的后果。
- 保证回答者可以跳过任何他们不愿意回答的问题。

如果研究设计允许的话,这些资料应该提前寄出或者直接送达回答者。无论如何,还是应该要求访谈员(若有访谈员)在开始访谈前再给回答者重复一下上述几点。

最后,也许有必要谈一下签署许可书的问题。一般来讲,我们不要求样本调查的回答者在完成一个调查前签署许可书。显然,在电话调查中,要求签署这样的许可书是行不通的。然而即使是在个人访谈调查里,大多数考虑周全的审查委员会都会认为要求回答者签署许可书的做法是不必要的,因为在大多数情况下,调查参与者所冒的风险都是很小的。另外,回答者在被问到每一个新问题时,都有重新考虑是否继续参加调查的

机会。

也有一些例外。在下列情况中,研究者和制度审查委员会很可能希望回答者以书面形式证明他们对调查有了足够的了解:

- 收集特别敏感的资料的时候。如果公开这些资料,可能会使回答者感到尴尬或使某些人受到伤害。
- 数据的保密性受到很大限制。
- 在对拒绝参与调查问题可能缺乏判断力或拒绝能力的总体进行调查时,如儿童、囚犯、雇员和学生等。
- 在使用以前收集的原来不是用于研究目的资料如试图用医疗记录作为调查答案的补充时。

上面提到的都是一些例外的情况,并非一般情况。只要保证在回答者同意参与之前能了解研究草案,大多数调查访谈不需要签署书面的同意书。无独有偶,希伯(Sieber,1992)对书面同意书的问题也曾进行过充分的讨论。

保护回答者

如果是从一个诸如保险计划成员或某机构雇员等的名单上抽取样本,必须坚持的一个最基本的道德原则是,不管样本成员是否同意参与调查,他们的生活在任何方面都不应该因此而受到不利的影响。也就是,样本成员是否回答都不应该被非研究团体的人知道。如果被抽取的样本成员不参加调查,也要保证他们不会有不利的结果(诸如他们的健康福利、服务、工作情况或者级别等)。

除此以外,有关保护调查回答者最主要的一点就是对他们所提供的资料的处理方式。与使用书面问卷或访谈内容明细表相比,直接向电脑输入答案一般较容易使数据具有保密性。为了减少泄密事情发生,细心的调查研究者往往会采取以下

步骤：

- 所有可以接触数据，或在数据收集中承担一定工作的人，都必须以书面形式对数据的保密性作出承诺。
- 尽量减少答案和身份识别符号之间的联系。名字或地址是最普通的身份识别符号。在姓名可以避而不用时，为了调查能顺利进行，通常都不要求在调查中使用姓名。很多调查机构都不在研究过程中的任何时候使用回答者的姓名。
- 在调查工具上有详细的身份识别符号时，如姓名或地址等，研究者应该把它们填入一种表格中以便随时都能把它们与实际的调查答案分开。研究者会尽快把身份识别符号（而不是识别符的编号）尽快从完成了的调查工具中去掉。
- 如果使用了名字或地址来识别样本或收集数据，在用完之后，研究者应该立即把它们删除或销毁。
- 允许非项目成员接触已经回收的调查数据。如果数据是书面形式的，最好把它们锁进某种文件柜。
- 不让那些可以从答案的特点猜测出回答者的人看到真实的调查回答（例如雇员调查里的管理人员或学生调查里的老师）。
- 实际的数据文件上通常每个回答者都会有自己的识别符编号。不允许数据文件的一般用户了解身份识别编号与样本地址或回答者身份识别符号之间的联系。
- 在数据分析期间，在公开一小类有可能被认出其身份的人的数据时，研究者要特别小心。
- 在项目完成后，或在调查工具的使用结束时，研究者要亲眼目睹完成了的调查研究工具的销毁，或确保其将继续被保存在安全之处。

显然，对于一个特定的研究项目来讲，它所需要采取的特定程序，可能与上述程序会有所不同。但是，上述程序所体现

的基本思路和关注点，理应在任何一个有责任心的研究者的项目中得到体现。[1]

回答者的酬劳

在大多数调查中，回答者获得的酬劳主要是精神上的，如他们在访谈过程中得到了享受或感觉他们做了值得做的事情。有时也会提供物质上的酬劳，例如，酬金、奖品，或服务。在给回答者提供酬劳时，必须注意，提供的方式不能给已经承诺的答案的保密性带来危害。除此以外，研究所应坚持的最重要的道德准则就是不夸大酬劳，并按许诺付酬。特别需要指出的是，酬劳固然能有一定作用，但是研究者争取得到合作的关键还是在于，能实事求是地介绍研究的用途，并承诺恰当的使用调查数据和不公开散发数据。

对访谈员负有的伦理责任

研究者要在两个方面对访谈员负责：第一，因为访谈员有责任把研究内容告知回答者，所以研究者的责任就是确保访谈员全面准确地了解有关研究的资料，不应该使访谈员处于一种被欺骗的、误导的或不适当的境地。第二，研究者必须解决访谈员的安全问题和对犯罪恐惧的问题。因为一般的住户样本

---

1. 法庭有可能向研究者索取问卷和记录，而研究人员可以以威胁到承诺的机密性为由保护问卷和记录。如果研究牵涉到特别敏感的材料，例如，司法研究中可能涉及的毒品或罪犯，研究者可以请求联邦或州机构保护问卷和记录不受传唤，或者在传票发出前，研究者可以销毁身份识别符号与答案间的联系。如果保持这个联系十分重要，例如一些纵向研究需要重复与回答者的联系，研究者要把保存有个人和答案之间的联系的文档送到国外或在美国法庭权限以外的地方。此外，希伯（Sieber，1992）的文章中提到了此问题更多的细节。

要涉及总体内的所有的地区，这样访谈员可能会去那些令他们感觉不安全的地方。在遇有这样的情况时，以下一些原则可能会对问题的妥善处理不无帮助：

- 在确定调查员尚未感到不安全之前，理应请他们驾车去看看样本所在地区。不仅各个邻里地区不尽相同，而且街区与街区之间也各有差别。
- 明确告诉访谈员，调查工作并未要求他们到那些他们感到不安全的地方进行调查。可供他们选择解决问题的方法包括：避免晚上做访问、尽量在周末去访问上班族、与其他访谈员一起访谈，或由有酬的护送员陪同进行访问等。让访谈员和现场督导一起，设计一种能令他们觉得安全的方式进行访谈也是一种不错的方法。
- 教给访谈员一些可减少自己受他人伤害的防身术。

幸运的是，访谈员真正出事的情况是非常少有的，但是社会里的确有犯罪现象存在，因此研究者和访谈员都需要感到访谈员已经清楚了去哪些地方或做哪些事情可能不安全，并且要使访谈员没有必须去那些地方和做那些事情的压力。

总　结

调查研究中的伦理问题与社会科学中的一般伦理问题没有什么不同。在大多数调查中，回答者（或访谈员）的真正的风险和可能的花费都很小，但仍然需要采取一些基本的步骤来减少参与者或社会科学研究者的风险。上面列出的步骤绝不是详尽无遗的，但是，不管是什么步骤，最重要的问题都是待人以诚。此外，我们还要坚持不懈地努力使利益最大化和使代价最小化，这些都是任何一个调查工作不可或缺的组成部分。

第 10 章　提供有关调查方法的信息

研究者在报告调查估计值时,在科学上有责任详细地报告使用的所有那些有可能对这些估计值产生影响的方法。此外,他们还应当计算自己报告的数字的准确性和精确性,并把计算结果公诸于众。本章讨论的内容,是一个全面的调查研究方法论的描述。

在研究者做出的诸多方法论决策中,完全错误的并不多见。的确存在着一些研究情境,在这些情境中,本书讨论到的任何折中方案对于资料收集来说是恰当和有成本效益的。然而对研究设计决策的评论是不应该脱离它的具体情境的,没有对数据收集方法做充分描述的做法都应该受到批评。让调查数据的读者和使用者都能够了解全面和完整的资料收集过程是十分重要的。

一个好的调查方法的描述一般有两种功能。第一种功能,是让人们了解到样本估计值怎样很好地描述了总体。只给出作者在这个方面的结论是不够的,而应该给出与精确度和偏差有关的详细运算,以便让读者做出自己的评估。第二种功能,是提供程序上的细节,以便其他研究者可以重复相同的资料收集过程,或找出可能会影响可比性的各调查间程序上的差异。

我们经常可以发现,有些调查报告只报告了一项调查的样本的量,而更认真的研究者,则在样本量之外,还会对抽样策略和回答率加以说明。尽管诸如这样的报告的合理的详细程度,会因数据使用的方式不同而有所不同,但是对任何调查都应该提供下列简明提纲所概括的信息:

- 抽样框(即从中抽取样本的那些人)及研究总体中那些有机会从抽样框中被抽取的人的比例的估计值和已知的那些被排除在样本之外的人与总体存在的总的差异。
- 抽样方法,包括任何对简单随机抽样的偏离,如整群、分层,或者总体子群的不等的选择率。
- 实际抽样结果,最初确定的样本处置方案:回答者的总数、无回答者的总数,以及不回答的主要原因。如果由于抽样框包括了不合格的单元(例如,非住户电话号码)导致不能精确计算回答率,研究者应该报告不合格的单元的数目和最可能的回答率估计值。美国舆论研究协会在2000年出版了一本论述报告回答率的专著,以提高报告的一致性以及研究者报告研究结果时所用术语的一致性。

- 被分析的问题的确切措辞：大型的调查报告应附有调查工具的完整复本。

除了对数据收集过程的事实加以说明以外，方法论附录还应包括其他五个方面的内容：

1. 由于很多报告不是提供给调查研究方法论专家看的，所以，有必要在方法论报告的引言部分，简要地介绍一下调查中可能出现的各种误差。

2. 报告应该报告由样本的独特设计引起的抽样误差的估计值。如果样本设计用的是分层法和整群法，或者抽取中使用了不同的概率，由这些设计特点引起的效应因不同的调查量度而异。一般讲，研究者应该计算其中的若干量度，包括他们预期受样本设计影响最大的和最小的量度的设计效应。然后，或仅报告这几个最受影响的题项的效应，或报告由这几个最受影响的题项的效应中概括出来的总效应的大致范围。

3. 提供有关访谈者可能会对数据产生影响的信息。这样的信息至少包括：访谈员总数；其中新培训的和经验丰富的访谈员各多少；访谈员接受的培训程度；是否对从他们的工作进行了抽样检查。如果访谈员的一些特征和调查内容有关，最好也提供他们的年龄、性别和种族等特征。

但是正如在第 7 章里提到的一样，如果分配给访谈员的回答者是基于方便的原则进行的，那么访谈员的影响是不能被可靠地计算出来的。格罗夫斯（Groves，1989）总结了很多有关访谈员的影响的资料，这些资料表明了访谈员对估计值影响的范围。如果能通过更多的研究得到评估访谈员影响的方法，并且将最后的评估值列为调查方法论报告的一项重要内容，应该是比较理想的。

4. 研究者应该尽可能地告诉读者无回答对样本估计值的影响。另外，如果研究者对原来未能获得他们的数据的人进行了抽样调查，也可将这样的调查结果公布出来。应鼓励访谈员

尽量从那些拒绝回答的人那里得到不回答的原因,以便研究者可以大致了解无回答者和回答者的区别。如果能够从其他来源得到有关总体的统计数据,例如相关的近期普查数据,那么研究者可以将样本与这些数据进行比较以估计无回答对样本的影响。

5. 一个好的方法论附录还应该包括一些有关调查中使用到的主要测量方法的信度和效度的数据。一般来说需要报告的相关资料有三种:

- 如果问题接受了认知实验室测试或系统的检测,那么报告中应该包括对这一事实及其结果的介绍。简单报告一下已经做过什么样的问题评估,对使用这个结果的人十分有益。值得报告的内容包括:问题被证明是可以理解的;前测中的行为编码表明问题可以按这样的措辞提问,并通常都可立即得到答案;有时前测表明某个问题还存在一定的问题,但出于某种原因我们仍然未作任何更正地保留了下来。显然,这些信息对使用从诸如这样的问题得到的结果的用户来讲是非常有用的。
- 研究者应提供评估答案效度所做的分析。这些答案与其他问题的答案在可预测的方式上的相关程度,可以证明这些问题所测量的东西是否是研究者所希望测量的东西。韦尔(Ware,1987)对全面评估信度和效度必须要做的各种分析做了一个很好的总结。
- 虽然在调查中很少能直接评估出事实数据报告的准确度,但是通过引用查核记录得来的类似项目的数据,可以给读者提供一个根据,使他们估计出调查估计值中的误差数量和方向。

关于问题能够在多大程度上测量出所希望测量的内容,到目前为止,还缺乏这个方面的系统性研究资料。但我们可以说,大多数调查报告提呈的都是面对效度(face validity)——答案的含义正是问题的设计者认为它们应该具有的含义。将以

评估问题和答案为目的的数据收集和分析变成调查方法论报告的例行部分,正是我们所希望的。

将这里列出的所有信息都包括在一个调查数据报告里,报告未免会过于详细。杂志通常都希望文章更简短一些。然而,所有这些信息显然都与调查质量的评估有关。在一个全面的调查分析报告中,完整的方法论附录应是不可少的。如果发表时需要缩短篇幅,至少也应该有关于数据收集过程细节的方法论报告。

总之,应该注意,收集和分析有关测量资料的过程很重要,对它们的报告同样很重要。显然,采取措施以使误差最小化是必要的,但是完全没有误差的调查是不可能的。报告测量是如何顺利进行的,以及评估结果中误差的数量和类别,是保证对调查资料合理使用的一个重要方面,也是构建进一步提高调查测量水平所依赖的知识的一个重要方面。

因此,在报告研究结论时,研究者有义务在报告中提供对调查数据的质量进行的评估,以及复制其得到的结果所需要的信息。要达到后一个目的必须仔细地描述所使用的调查程序,而实现前一个目标需要特别重视误差测量和结果的传达。在一个报告对诸如问题是否清晰或访谈员是否影响到答案等有关误差问题避而不谈时,大多数读者会认为这个报告是没有问题的。研究者至少能让读者知道影响调查估计值误差的各种不同来源(这是首要的)。但是,从长远来看,在报告调查结果时,将那些为提供详细的误差估计值所需要的步骤包括进去,应成为一种惯例。

练　习

使用本章提到的标准,对建立在调查基础上的一本著作或一个研究报告的方法论部分的恰当性和完整性进行全面系统的评价。

第 11 章　调查误差总览

调查的整体设计涉及周密地考虑一个调查的所有方面,和为一个特定的研究目标选择适当的设计标准。本章将讨论优秀的调查设计和实践中最常见的偏差,同时还要评估优秀调查设计和实践的成本节俭的可能性以及它们对调查估计值的精确度、准确度和信度的意义。

整体调查设计的概念

整体调查设计意味着在设计调查或评估调查数据质量时，研究者要着眼于整个的数据收集过程，而不只是简单地考虑调查的某一两个方面。样本（抽样框、样本量、样本设计和回答率）的质量、作为测量工具的问题的质量、数据收集（特别是有效的访谈员培训和督导措施的使用）的质量以及收集的形式构成了一系列紧密联系的问题和设计策略，所有这些方面都有影响最终调查结果的可能性。因此对调查整体设计的全面正确评价应有三个具体的含义：

- 在设计一个调查数据收集程序时，研究者要自觉地考虑到保持成本和在调查设计过程中各个方面的方法论的严密性之间的平衡。如果调查的其他方面不能保证得到投资于消减调查的某一各方面的误差的投资的同样水平的投资，用于该一方面的投资则不会奏效。
- 在评估数据质量时，研究者应当询问所有影响数据质量的决策是如何做出和实行的。
- 在报告调查的细节时，研究者应该比较详细地报告影响数据误差水平的数据收集工作的所有方面。

误差总览

对本书讨论到的方法论选择的意义进行归纳以适用于各种不同的研究是难以做到的。因为严谨决策的成本在不同的研究情景中差异巨大，在不同的方法选择之间也存在巨大差异。同样，为做出省钱的决策而导致错误的可能性也有非常大的不同。为一个调查制定或选择设计方案是指研究者应仔细

评估各种设计选择，以及评估其成本和误差出现的概率，然后决定选用何种设计。

给研究者想要研究的每一个总体成员一个已知的选择机会的完美的抽样框几乎是不存在的，因此，对谁真正可以成为样本这一决策的评估是不能脱离具体情境而进行评估的。尽管如此，研究者仍然有责任弄清其抽样框的完整性以及漏掉了哪些人，并且不能暗示其样本估计值也适合于那些没有机会被抽样到的人。

在调查研究中最为普遍存在的问题是节约成本的问题，它最可能发生在抽样领域。其中最糟糕的情况是，研究者试图从那些完全没有用概率抽样方法抽到的人群中收集数据以总结出结论。例如，从自愿填答期刊上调查问卷的那些读者那里收集数据。然而，对于一些很难获得合作的样本，通常的解决办法只能是使用非概率样本组成的更易得和更易取得配合的人群。这一方面比较典型是很多民意、政治和市场研究测验中进行的调查。事实上，对于电话调查来说，样本替代不会节约多少钱，尽管它能让调查迅速地进行。而对于面对面的家庭访谈，在概率和非概率样本之间成本就会有许多差异。为了节约这种成本付出的主要代价是放弃有关数据误差的统计数据的信度。并没有描述样本和样本抽取的总体之间关系的科学根据。如果我们不能满足于手头上掌握的总体的来自其他途径的一些信息，而想得到一些有关总的更为广泛的信息，那么这种非统计学的抽样程序可能会更好地满足这种目的要求。然而，如果我们的兴趣不限于量的估计值的次序，且科学可靠性问题也在必须考虑的问题之列时，那么对样本进行替代所节约的成本可能会是得不偿失的。

调查中另一个几乎同样普遍的问题是接受低回答率。几乎所有关于邮件调查中早期返回的问卷的研究都表明，早期返回的问卷是一个有偏倚的样本，这种偏倚的存在直接影响着调查的主题。电话调查中与低回答率有关的偏倚没那么明显，尽管它总是很可能将那些教育程度低的群体、老年群体，及那些

具有独特观点的群体排除在调查之外。真正的问题在于没有准确的信息可以判断无回答何时会严重影响估计值,因此很难断定说回答率何时真正算低,因而必须做一些额外的补救工作。研究者所做工作的最低限度是要保证可靠性,因此基于低回答率的调查数据很容易遭受非难,研究者不得不确定其影响有多大。

由于大多数考虑调查设计的研究者会考虑到抽样,所以这里就不花更多的时间讨论这个问题。如果要选取精密复杂的样本,就需要合格的抽样统计学家的帮助,因为只有这样的统计学家才能考虑到计算抽样误差的不同抽样方案之间的平衡。这里应该指出的一点是,总体样本的大多数都牵涉到整群抽样。事实上,多阶整群样本是处理大范围抽样问题的最好办法。但是,经常可以发现,在一些报告中整群抽样的影响在评估抽样误差时被忽略了。如果研究者没有一个有效的简单随机样本,而实际情况也常常如此,那么在评估标准误差时,他们就不能使用简单随机抽样假设。

数据收集模式的选择是影响调查成本的最重要的选择之一。虽然多年来,个人访谈调查的各种方法曾被认为是进行一般总体调查的唯一有效的方法,但是现在在大多数调查机构中使用电话调查的方法比个人访谈调查更为普遍。

电话访谈已被证明是一种产生的调查数据能满足很多目的的有效途径。但是,电话调查的成本往往使得这一方法并不能在所有的情况下都能成为我们的首选。依据具体情况的不同,电话访谈可能会存在一些不足:排除了总体中没有电话的部分(抽样框的问题);会产生较高的无回答率(虽然这并不常发生);会使研究者收集的数据更少(因为电话访谈的时间通常比个人访谈的时间短)。但总的来说,对于很多项目来讲,电话调查方法的确是一种正确选择。与大多数节约成本的调查方法一样,在有些情况下,选择电话调查的非货币代价可能会很高。

就某些总体的数据收集来说,邮件和互联网显然是有诱惑

力的方式。由于其相对低廉的价格,邮件和互联网常被用来与各种花费高的调查方法结合使用:用廉价的邮件和互联网方式从使用邮件和互联网的人群中收集数据;那些不太愿意回答问题的人或者不便于用邮件和互联网方式回答问题的人(例如,因为他们很少接触互联网)的数据,则用其他的数据收集方式收集。优秀的调查设计要能以省钱高效并获得高回答率的方式解决数据收集方式与样本、研究主题、工具需求之间的匹配关系问题,而糟糕的调查设计则只依靠一种对研究总体的某些部分不合适或效率低的方式收集数据。

访谈员的质量也是调查研究里最易忽略的问题。事实证明,良好的培训和监督可以对访谈员进行的访谈产生巨大作用。例如,对于大多数一般的调查问题来说,没有经过良好培训或监督的访谈员可能会增加围绕估计值误差的幅度,就像取了一个更小的样本一样。这就是说如果存在访谈员对答案的严重影响,一个 1 000 人的样本产生的样本精确度可能只及一个 700 人的样本。虽然访谈质量的重要性随调查的内容和问题类型的不同而不同,但是在大多数一般目的的调查中,都有部分问题受到了访谈员的严重影响。在这情况下,削减对访谈员的培训和监督的费用就是一个错误的决定。

另外,访谈人员的素质也会影响到回答率。使用能够取得回答者合作的优秀访谈员,是保证高回答率的最简单的一种方法。因此需要对访谈员严格督导、再培训,并排除那些不善于取得合作的访谈员,以便减少无回答率和可能出现的额外支出。

将访谈员向回答者解释访谈任务的方式格式化为一些标准化程序,可以用于调查研究。事实证明,这种技术是一种实际上不需成本而能提高回答者回答问题的水平的方式。

最后,对调查问题全面的评估可能是最重要和最划算的减少调查误差的一种方法。例如:通过对焦点小组和认知访谈的合理运用,研究者可以提高回答者对问题理解的一致性和答案的有效性。而在实地前测中对行为的系统研究可以鉴别出访

谈员是否以一致的方式提出问题或把问题展现给回答者。提高问题的质量既可以提高回答者报告的效度又可以减少访谈员对答案的影响。

总　结

优秀的设计和实践的目的是让每一块钱能够最大可能地获得准确性、可靠性和可重复性。而一个特定研究的精确性、有效性和可重复性是由已经确认的问题的性质和资料的使用方式决定的。

人们有时会读到一些文章，它们认为社会科学是不精确的，并且还对社会科学和物理科学的测量过程做过引起人们反感的比较。尽管这种见解是常见的，但它们通常也是无知的。物理科学和生物医学中的一些基本测量，诸如血液样本中铅的含量、血压的读数、X 光的读数和金属弹性的测量等，都被证明具有较低的可信度。这些领域里的测量都可能产生可好可坏的结果，这取决于使用什么样的方法论以及设计测量过程谨慎与否，这个方面特纳和马丁（Turner & Martin，1984）提供了很多例子。对于调查研究来说，情况也是一样。

使调查里误差最小化，一般是需要付出一定的代价的。在有些情况下，最好的测量，其价值却可能抵不上付出的代价。通过对家庭进行抽样来抽取总体的样本，通常会排除那些与家庭没有关系的人。在大多数地区，这部分与家庭无关的人只占总体中很小的一部分，但是要设计一种方法使样本中也能包括这些与家庭无关的人，其花费会是很高的。

无回答率可以减少到零。人口普查局在进行国家健康访谈调查时，回答率超过了 90%，只是在一些主要的大城市回答率要低一些。而且如果有人想投入足够的时间和金钱，即使是在大城市的最不好客的街区，其回答率仍可能会接近 100%。然而也有花费极高，潜在误差减少却很有限的情况。

幸运的是，回答者能够并且愿意回答社会科学家和决策者感兴趣的很多问题。但是对于另外一些问题，假如回答者能够并且愿意在标准调查过程中回答这些问题，就会给研究者带来很大的方便，可是回答者往往却不这样做。例如，使用标准调查技术调查酒后驾车和破产问题时，这些情况被回答者明显低报了（Locander，Sudman，& Bradburn，1976）。或许一项研究可以拥有某种方法，以引导回答者准确地报告这些事实，但这样需要研究者付出比通常更大的努力来赢得回答者的合作。研究者还得依据将来使用数据的方式，对数据的准确度和详细度问题做出决策。读者可以参看格罗夫斯（Groves，1989）的文章，其中对调查成本和调查误差之间的关系做了更为广泛的分析。

人们一直认为，在一个调查的情景中，人们能够且愿意告诉我们的东西是有一定限度的。然而，这些限度有被夸大之嫌。可用标准化调查程序精确地量度的东西，确实是有一定限度的。但是这些限度可能更多地与预算和研究者在测量过程中所付出的努力的多少有关，而不是与标准化调查是否切实可行有关。

研究者必须在具体的情景下来评价一个设计。一个设计的种种选择究竟是否正确，要看它们是否明智、是否会形成符合预期目的的数据。必须指出的是，调查误差并不仅仅来源于决策的严密性和成本节约两个方面，还来源于对无回答率影响的评价、问题设计以及访谈员行为对调查估计值质量的影响。由于我们对数据收集这些方面的问题缺乏足够的重视，因而经常会得到缺乏信度或不符合信度的最新标准的调查数据（例如参看 Bailar & Lanphier，1978；Turner & Martin，1984）。

本书所推崇的整体调查概念包括以下几个方面：

- 在数据收集上不存在任何妨碍研究者使用数据来达到调查目的的弱点。
- 对资料收集的各个方面的设计要保持相对的一致，以便避

免出现对某一方面的投入提高了精确度却无法被其他方面的精确度的变化证明是合理的情况。

- 调查数据的用户在使用基于一个样本调查的估计值的时候,能对调查中的可能的误差源、数据的精确程度和基于调查估计值的置信度等予以适当考虑。

最后,希望研究的使用者对自己提出的关于任何调查中数据收集的那些问题有一个全面的把握;希望研究者对设计决策细节的重要性有更好的把握;希望所有的读者可以用新学到的知识致力于更优秀的调查设计和实施。

参考文献

American Association for Public Opinion Research. (2000). *Standard definitions: Final dispositions of case codes and outcome rates for surveys.* Ann Arbor, MI: AAPOR.

Anderson, B., Silver, B., & Abramson, P. (1988). The effects of race of the interviewer on measures of electoral participation by blacks. *Public Opinion Quarterly*, 52(1), 53-83.

Aquilino, W. S. (1994). Interview mode effects in drug surveys. *Public Opinion Quarterly*, 58(2), 210-240.

Aquilino, W. S. (1998). Effects of interview mode on measuring depression in younger adults. *Journal of Official Statistics*, 14(1), 15-30.

Aquilino, W. S., & Losciuto, L. A. (1990). Effects of mode of interview on self-reported drug use. *Public Opinion Quarterly*, 54(3), 362-391.

Bailar, B., & Lanphier, C. (1978). *Development of survey research methods to assess survey practices.* Washington, DC: American Statistical Association.

Baker, R. P., & Lefes, W. L. (1988). The design of CATI systems: A review of current practice. In R. M. Groves, P. N. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nichols II, & J. Waksberg (Eds.), *Telephone survey methodology.* New York: John Wiley.

Beatty, P. (1995). Understanding the standardized/non-standardized interviewing controversy. *Journal of Official Statistics*, 11 (2), 147-160.

Belson, W. A. (1981). *The design and understanding of survey questions.* Aldershot, England: Gower.

Berry, S., & Kanouse, D. (1987). Physician response to a mailed survey: An experiment in timing of payment. *Public Opinion Quarterly*, 51(1), 102-114.

Billiet, J., & Loosveldt, G. (1988). Interviewer training and quality of responses. *Public*

Opinion Quarterly, 52(2), 190-211.

Bishop, G. F., Hippler, H. J., Schwartz, N., & Strack, F. (1988). A comparison of response effects in self-administered and telephone surveys. In R. M. Groves, P. N. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nichols II, & J, Waksberg (Eds.), *Telephone survey methodology* (pp. 321-340). New York: John Wiley.

Blom, E., & Lyberg, L. (1998). Scanning and optical character recognition in survey organizations. In M. Couper et al. (Eds.), *Computer-assisted information collection* (pp. 449-520). New York: John Wiley.

Bradburn, N. M., & Sudman, S. (1992). The current status of questionnaire design. In P. N. Beimer, R. M. Groves, L. E. Lyberg, N. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 29-40). New York: John Wiley.

Bradburn, N. M., Sudman, S., & Associates. (1979). *Improving interview method and questionnaire design.* San Francisco: Jossey-Bass.

Brick, J. M., Waksberg, J., Kulp, D., & Starer, A. (1995). Bias in list-assisted telephone samples. *Public Opinion Quarterly*, 59, 218-235.

Bryson, M. (1976, November). The Literary Digest poll: Making of a statistical myth. *American Statistician*, 184-185.

Burton, S., & Blair, E. (1991). Task conditions, response formulation processes, and response accuracy for behavioral frequency questions in surveys. *Public Opinion Quarterly*, 55, 50-79.

Cannell, C., & Fowler, F. (1964). A note on interviewer effect in self-enumerative procedures. *American Sociological Review*, 29, 276.

Cannell, C., Groves, R., Magilavy, L., Mathiowetz, N., & Miller, P. (1987). An experimental comparison of telephone and personal health interview surveys. *Vital & Health Statistics*(Series 2, No. 106). Washington, DC: Government Printing Office.

Cannell, C., Marquis, K., & Laurent, A. (1977). A summary of studies. *Vital & Health Statistics* (Series 2, No. 69). Washington, DC: Government Printing Office.

Cannell, C., & Oksenberg, L. (1988). Observation of behaviour in telephone interviewers. In R. M. Groves, P. N. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nichols II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 475-495). New York: John Wiley.

Cannell, C., Oksenberg, L., & Converse, J. (1977). *Experiments in interviewing techniques: Field experiments in health reporting*, 1971-1977. Hyattsville, MD: National Center for Health Services Research.

Catania, J. A., Gibson, D., Chitwood, D., & Coates, T. (1990). Methodological prob-

lems in AIDS behavioral research: Influences on measurement error and participation bias in studies of sexual behavior. *Psychological Bulletin*, 108(3), 339-362.

Catlin, G., & Ingram, S. (1988). The effects of CATI on costs and data quality. In R. M. Groves, P. N. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nichols II, & J. Waksberg (Eds.), *Telephone survey methodology*. New York: John Wiley.

Conrad, E G., & Schober, M. E (2000). Clarifying question meaning in a household telephone survey. *Public Opinion Quarterly*, 64(1), 1-28.

Converse, J. (1987). *Survey research in the United States*. Berkeley: University of California Press.

Converse, J., & Presser, S. (1986). *Survey questions*. Beverly Hills, CA: Sage.

Couper, M. P. (1999). The application of cognitive science through computer assisted interviewing. In M. G. Sirken et al. (Eds.), *Cognition in survey research* (pp. 277-300). New York: John Wiley.

Couper, M. P., Barkin, R. P., Bethlehem, J., Clark, C. E, Martin, J., Nicholls, W. L., & O'Reilly, J. M. (Eds.). (1998). *Computer assisted survey information collection*. New York: John Wiley.

Couper, M. P., Hansen, S. E., & Sadowsky, S. A. (1997). Evaluating interviewer use of CAPI technology. In L. E. Lyberg, P. Beimer, M. Collins, et al. (Eds.), *Survey measurement and process quality* (pp. 267-286). New York: John Wiley.

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychiatrika*, 16, 297-334.

Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.

de Leeuw, E. D., & van der Zouwen, J. (1988). Data quality in telephone and face to face surveys: A comparative meta-analysis. In R. M. Groves, P. N. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nichols II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 283-299). New York: John Wiley.

DeMaio, T. J., & Rothgeb, J. M. (1996). Cognitive interviewing techniques: In the lab and in the field. In N. A. Schwarz & S. Sudman (Eds.), *Answering questions* (pp. 177-196). San Francisco: Jossey-Bass.

Department of Health and Human Services. (1996). *Protection of human subjects* (Title 45, Code of Regulations, Part 46). Washington, DC: Government Printing Office.

DeVellis, R. E (1991). *Scale development: Theory and applications*. Newbury Park, CA: Sage.

Dielman, L., & Couper, M. P. (1995). Data quality in CAPI survey: Keying errors. *Journal of Official Statistics*, 11(2), 141-146.

Dillman, D. A. (2000). *Mail and Intemet surveys: The tailored design method.* New York: John Wiley.

Dillman, D. A., Carpenter, E., Christensen, J., & Brooks, R. (1974). Increasing mail questionnaire response: A four state comparison. *American Sociological Review*, 39(5), 744-756.

Dillman, D. A., & Miller, K. J. (1998). Response rates, data and feasibility for optical scannable mail surveys for small research centers. In M. P. Couper et al. (Eds.), *Computer assisted information collection* (pp. 475-498). New York: John Wiley.

Dillman, D. A., & Tarnai, J. (1991). Mode effects of cognitively designed recall questions: A comparison of answers to telephone and mail surveys. In P. N. Beimer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 367-393). New York: John Wiley.

Dykema, J., Lepkowski, J. M., & Blixt, S. (1997). The effect of interviewer and respondent behavior on data quality: Analysis of interaction coding in a validation study. In L. Lyberg, P. Beimer, M. Collins, et al. (Eds.), *Survey measurement and process quality* (pp. 287-310). New York: John Wiley.

Edwards, W. S., et al. (1 994). Evaluation of National Health Interview Survey diagnostic reporting. In *Vital & Health Statistics* (Series 2, No. 120). Hyattsville, MD: National Center for Health Statistics.

Edwards, W. S., Winn, D. M., & Collins, J. G. (1996). Evaluation of 2-week doctor visit reporting in the National Health Interview Survey. In *Vital & Health Statistics* (Series 2, No. 122). Hyattsville, MD: National Center for Health Statistics.

Erlich, J., & Riesman, D. (1961). Age and authority in the interview. *Public Opinion Quarterly*, 24, 99-114.

Filion, F. (1975). Estimating bias due to nonresponse in mail surveys. *Public Opinion Quarterly*, 39(4), 482-492.

Forsyth, B. H., & Lessler, J. T. (1992). Cognitive laboratory methods: A taxonomy. In P. N. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 393-418). New York: John Wiley.

Fowler, F. J. (1991). Reducing interviewer related error through interviewer training, supervision, and other means. In P. N. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 259-278). New York:

John Wiley.

Fowler, F. J. (1992). How unclear terms affect survey data. *Public Opinion Quarterly*, 56(2), 218-231.

Fowler, F. J. (1995). *Improving survey questions.* Thousand Oaks, CA: Sage.

Fowler, F. J. (1998). Nonresponse bias in mail surveys of health plan members. In 1998 *proceedings*, *Section on Survey Research Methods*, *American Statistical Association*, pp. 576-580.

Fowler, F. J., & Cannell, C. F. (1996). Using behavioral coding to identify cognitive problems with survey questions. In N. A. Schwarz & S. Sudman (Eds.), *Answering questions* (pp. 15-36). San Francisco: Jossey-Bass.

Fowler, F. J., & Mangione, T. W. (1990). *Standardized survey interviewing: Minimizing interviewer related error.* Newbury Park, CA: Sage.

Fowler, F. J., Roman, A. M., & DJ, Z. X. (1998). Mode effects in a survey of Medicare prostate surgery patients. *Public Opinion Quarterly*, 62(1), 29-46.

Fox, J. A., & Tracy, P. E. (1986). *Randomized response: A method for sensitive surveys.* Beverly Hills, CA: Sage.

Fox, R., Crask, M., & Kim, J. (1988). Mail survey response rate: A meta-analysis of selected techniques for increasing response. *Public Opinion Quarterly*, 52(4), 467-491.

Friedman, P. A. (1942). A second experiment in interviewer bias. *Sociometry*, 15, 378-381.

Groves, R. M. (1989). *Survey errors and survey costs.* New York: John Wiley.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys.* New York: John Wiley.

Groves, R. M., & Kahn, R. L. (1979). *Surveys by telephone.* New York: Academic Press.

Groves, R. M., & Lyberg, L. (1988). An overview of nonresponse issues in telephone surveys. In R. M. Groves, P. N. Biemer, L. E. Lyberg, J. T Massey, W. L. Nichols II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 191-212). New York: John Wiley.

Heberlein, T., & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *American Sociological Review*, 43, 447-462.

Henry, G. T. (1990). *Practical sampling.* Newbury Park, CA: Sage.

Henson, R., Roth, A., & Cannell, C. (1977). Personal versus telephone interviews: The effect of telephone reinterviews on reporting of psychiatric symptomatology. In C. Cannell,

L. Oksenberg, & J. Converse (Eds.), *Experiments in interviewing techniques: Field experiments in health reporting*, 1971-1977. Hyattsville, MD: National Center for Health Services Research.

Hochstim, J. (1967, September). A critical comparison of three strategies of collecting data from households. *Journal of the American Statistical Association*, 62, 976-989.

Hyman, H., Feldman, J., & Stember, C. (1954). *Interviewing in social research.* Chicago: University of Chicago Press.

Jabine, T B., Straf, M. L., Tanur, J. M., & Tourangeau, R. (Eds.). (1984). *Cognitive aspects of survey methodology: Building a bridge between disciplines.* Washington, DC: National Academy Press.

James, J., & Bolstein, R. (1990). The effect of monetary incentives and follow-up mailings on the response rate and the response quality in mail surveys. *Public Opinion Quarterly*, 54 (3), 346-361.

Jobber, D. (1984). Response bias in mail surveys: Further evidence. *Psychological Reports*, 54, 981-984.

Kalton, G. (1983). *Introduction to survey sampling.* Beverly Hills, CA: Sage.

Keeter, S., Miller, C., Kohut, A., Groves, R. M., & Presser, S. (2000). Consequences of reducing nonresponse in a national telephone survey. *Public Opinion Quarterly*, 64(2), 125-148.

Kish, L. (1949). A procedure for objective respondent selection within the household. *Journal of the American Statistical Association*, 44, 380-387.

Kish, L. (1962). Studies of interviewer variance for attitudinal variables. *Journal of the American Statistical Association*, 57, 92-115.

Kish, L. (1965). *Survey sampling.* New York: John Wiley.

Lepkowski, J. M. (1988). Telephone sampling methods in the United States. In R. M. Groves, P. N. Biemer, L. E. Lyberg, J. T Massey, W. L. Nichols II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 73-98). New York: John Wiley.

Lessler, J. T., & Forsyth, B. H. (1996). A coding system for appraising questionnaires. In N. A. Schwartz & S. Sudman (Eds.), *Answering questions* (pp. 259-292). San Francisco: Jossey-Bass.

Lessler, J., & Tourangeau, R. (1989, May). Questionnaire design in the cognitive research laboratory. *Vital & Health Statistics* (Series 6, No. 1). Washington, DC: Government Printing Office.

Likert, R. (1932). A technique for measurement of attitudes. *Archives of Psychology*, 140.

Little, R., & Rubin, D. (1987). *Statistical analysis with missing data.* New York: John Wiley.

Locander, W., Sudman, S., & Bradburn, N. (1976). An investigation of interview method, threat and response distortion. *Journal of the American Statistical Association*, 71, 269-275.

Lohr, S. L. (1998). *Sampling design and analysis.* New York: Brooks-Cole.

Mangione, T., Hingson, R., & Barter, J. (1982). Collecting sensitive data: A comparison of three survey strategies. *Sociological Methods and Research*, 10(3), 337-346.

Marin, G., & Marin, B. V. (1991). *Research with Hispanic populations.* Newbury Park, CA: Sage.

Marquis, K., Cannell, C., & Laurent, A. (1972). Reporting health events in household interviews: Effects of reinforcement, question length, and reinterviews. *Vital & Health Statistics* (Series 2, No. 45). Washington, DC: Government Printing Office.

McDowell, I., & Newell, C. (1996). *Measuring health: A guide to rating scales and questionnaires.* Oxford: Oxford University Press.

McHorney, C. A., Kosinski, M., & Ware, J. E. (1994). Comparisons of the cost and quality of norms for the SF-36 Health Survey by mail versus phone interviews: Results from a national survey. *Medical Care*, 32(6), 551-567.

Miller, P. V., & Cannell, C. F. (1977). Communicating measurement objectives in the interview. In P. M. Hirsch, F. G. Kline, & P. V. Miller (Eds.), *Strategies for communication research.* Beverly Hills, CA: Sage.

Nichols, W. L., II. (1988). Computer-assisted telephone interviewing: A general introduction. In R. M. Groves, P. N. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nichols II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 337-385). New York: John Wiley.

Nichols, W. L., Baker, R. P., & Martin, J. (1997). The effect of new data collection technologies on survey data quality. In L. Lyberg et al. (Eds.), *Survey measurement and process quality* (pp. 221-248). New York: John Wiley.

Oksenberg, L., Cannell, C., & Kalton, G. (1991). New strategies of pretesting survey questions. *Journal of Official Statistics*, 7(3), 349-366.

Payne, S. (1951). *The art of asking questions.* Princeton, NJ: Princeton University Press.

Penne, M. A., Lessler, J. T., Beiler, G., & Caspar, R. (1998). Effects of experimental and audio computer-assisted self-interviewing (ACASI) on reported drug use in the HSDA. In 1998 *proceedings, Section on Survey Research Methods, American Statistical Association*,

pp. 744-749.

Robinson, D. , & Rhode, S. (1946). Two experiments with an anti-Semitism poll. *Journal of Abnormal Psychology*, 41, 136-144.

Robinson, J. P. (1989). Poll review: Survey organization differences in estimating public participation in the arts. *Public Opinion Quarterly*, 53(3), 397-414.

Robinson, J. P. , Shaver, P. R. , & Wrightsman, L. S. (Eds.). (1997). *Measures of personality and social psychological attitudes* (Vol. 1). New York: Harcourt Brace.

Royston, P. N. (1989). Using intensive interviews to evaluate questions. In F. Fowler (Ed.), *Conference proceedings: Health survey research methods* (pp. 3-8). Washington, DC: National Center for Health Services Research.

Saris, W. (1991). *Computer-assisted interviewing.* Newbury Park, CA: Sage.

Schaeffer, N. C. (1992). Interview: Conversation with a purpose or conversation? In P. N. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 367-393). New York: John Wiley.

Schober, M. F. , & Conrad, F. G. (1997). Conversational interviewing. *Public Opinion Quarterly*, 61(4), 576-602.

Schuman, H. , & Converse, J. (1971). The effects of black and white interviewers on black responses in 1968. *Public Opinion Quarterly*, 35, 44-68.

Schuman, H. , & Presser, S. (1981). *Questions and answers in attitude surveys.* New York: Academic Press.

Schwartz, N. A. , & Sudman, S. (Eds.). (1996). *Answering questions.* San Francisco: Jossey-Bass.

Sieber, J. (1992). *Planning ethically responsible research: Developing an effective protocol.* Newbury Park, CA: Sage.

Singer, E. , Van Hoewyk, J. , Gebler, N. , Raghunnathan, T, & McGanagle, K. (1999). The effect of incentives on response rates in interviewer-mediated surveys. *Journal of Official Statistics*, 15(2), 217-230.

Singer, E. , Van Hoewyk, J. , & Maher, M. P. (2000). Experiments with incentives in telephone surveys. *Public Opinion Quarterly*, 64(2), 171-188.

Sirken, M. G. , et al. (Eds.). (1999). *Cognition in survey research.* New York: John Wiley.

Suchman, L. , & Jordan, B. (1990). Interactional troubles in face-to-face survey interviews. *Journal of the American Statistical Association*, 85, 232-241.

Sudman, S. (1967). *Reducing the cost of surveys.* Chicago: Aldine.

Sudman, S. (1976). *Applied sampling.* New York: Academic Press.

Sudman, S., & Bradburn, N. (1982). *Asking questions.* San Francisco: Jossey-Bass.

Tanur, J. M. (1991). *Questions about questions.* New York: Russell Sage.

Thornberry, O. T, & Massey, J. T. (1988). Trends in United States telephone coverage across time and subgroups. In R. M. Groves, P. N. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nichols II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 25-50). New York: John Wiley.

Thnrstone, L., &Chave, E. (1929). *The measurement of attitude.* Chicago: University of Chicago Press.

Tourangeau, R., & Smith, T. W. (1998). Collecting sensitive data with different modes of data collection. In M. P. Couper et al. (Eds.), *Computer assisted survey information collection* (pp. 431-453). New York: John Wiley.

Turner, C. F., Forsyth, B. H., & O'Reilly, J. M., et al. (1998). Automated self-interviewing and the survey measurement of sensitive behaviors. In M. P. Couper et al. (Eds.), *Computer assisted survey information collection* (pp. 455-473). New York: John Wiley.

Turner, C. F., & Martin, E. (1984). *Surveying subjective phenomena.* New York: Russell Sage.

Waksberg, J. (1978). Sampling methods for random-digit dialing. *Journal of the American Statistical Association*, 73, 40-46.

Ware, J. (1987). Standards for validating health measures: Definition and content. *Journal of Chronic Diseases*, 40, 473-480.

Warriner, K., Goyder, J., Jersten, H., Hohner, P., & McSpurren, K. (1996). Cash versus lotteries versus charities in mail surveys. *Public Opinion Quarterly*, 60(4), 542-562.

Weiss, C. (1968). Validity of welfare mothers' interview responses. *Public Opinion Quarterly*, 32(4), 622-633.

Willis, G. B., DeMaio, T., & Harris-Kojetin, B. (1999). Is the bandwagon headed to the methodicalogical promised land? Evaluating the validity of cognitive interviewing techniques. In M. G. Sirken et al. (Eds.), *Cognition in survey research* (pp. 133-154). New York: John Wiley.

附录

从调查误差的来源及其有效规避来看总体研究设计

——福勒《调查研究方法》简评

◆舒建军[1]

在进行具体的社会学或者更广义的社会科学方面的研究之前，对研究主题进行认识论和方法论的推敲是十分必要的。虽然在认识论和方法论方面，西方社会科学界的讨论源远流长，但在中国内地，这方面确实存在不少问题。虽然有这方面的课程和教学，但知识论和方法论的锤炼并没有成为学者的自觉意识，往往同研究主题隔离开了，变成了一个点缀或者装饰。把认识论和方法论贯穿到研究过程之中，同具体的研究融为一体，这不仅能使研究更为完整和全面，从学科意义上讲，别人知道你是怎么开展研究或者在哪些具体方法支撑下开展研究的就能大致了解你在某个问题上有何推进，因此，这种融合还能切实提升人类的增量知识，而不是目前所看到的，重复研究、自说自话等。

在从事社会调查和研究的时候，一般都知道有些什么方法进行资料的收集和整理、分析。但到具体应用时，不免盲人骑瞎马，乱点鸳鸯谱。明明相关性不强，硬要拉上一两个常用的

---

1 舒建军，中国社会科学院社会学所博士研究生，现供职于《中国社会科学》杂志社。早期的学科训练是世界历史，专门探讨了西方中心论单一现代性叙事下的英国或者西欧近代早期的另类发展模式。现在对于非地理意义上的乡土中国问题思考得多一些。近年主要的撰述有：《待清偿与清偿过的权利》(《读书》2001 年)、《公共权利与当代中国社会》(《社会科学论坛》2004 年)、《清开“三农”壅塞的言路》(《开放时代》2004 年)等。

方法充门面。更多的问题是对各研究方法应用的限度不甚了解,结果造成相当多的误差。福勒的《调查研究方法》一书很有针对性,对我们这些还刚刚入门的人来说是一个很及时和充分的提醒:总体研究设计中,从问题设计到开展调查直到整理资料的每一个环节都存在影响调查结论真实性的各种因素,认识到这些问题并有效规避,不仅仅是纯粹的技巧,它还是完善各种研究方法的题中应有之义。

福勒在开篇就直呈自己有这种意图,本书"将在误差产生的原因和调查数据收集在方法论上所涉及的领域,给读者一个系统而全面的认识。"

抽样误差是总体调查误差的一个重要方面。首先,福勒强调评估样本的方式不是经由结论和样本本身的特点,而是通过考查样本选择的过程。在选择样本的过程中,应注意评估抽样框的三个原则:全面性、概率的度量和有效性。在抽取一阶样本和多阶样本的过程中,不仅要注意每种抽样策略的程序,而且要注意这些抽样样本所能产生的抽样误差。在样本规模上的一些误解也得到了澄清:第一种误解过度关注样本中包含的总体的分值大小与样本规模之间的关系,福勒指出分值对抽样误差的影响微不足道;第二种误解是受了"标准调查"的影响,据此来确定自己的样本规模,这种参照显然没有从自己所要达到的研究目标和研究设计的具体方面来考虑样本规模;第三种误解同第二种有些类似,只不过看似更为科学一些,即先确定允许范围内的最大误差值或者所要求的估计值的精确度,再根据抽样变异导致的不同置信区间或者类似方法计算出保证此一精确度的样本规模,这种方法有点投机取巧,但在实际情形下行不通。决定样本规模的主要因素还是具体的分析规划,这个规划的关键部分不是估计值,而是那些重要的由一些反映总体的参数形成的子群。另外,非抽样误差与抽样误差一样需要同等对待。

在样本设计的实施过程中会出现不应答的情况。这种无回应对调查估计值会产生影响并直接使调查结果发生偏误。

与无回应有关的偏误包括调查对象的不可得性和不可接近性，这方面的具体例子很多。福勒探讨了如何减少各种调查中的无回应率的方法。在电话或当面访谈调查中为了减少由于联系不上而产生的无回应率，应集中在晚上和周末多打电话并为访谈对象提供弹性进度表以方便后续访谈。当然事先提供的材料也非常重要，例如，邮寄预告函和研究项目的关键信息并确保该项调查本身及其方式不会对访谈对象产生不利影响。另外，访谈员的素质也非常重要。为了减少邮件调查中的无回应率，可做的事情有很多，例如可采取各种措施使调查问卷看上去更专业、更人性化、更引人注目，同时设计更好的自填问卷以及保持不间断的联系等。有些无回应是不可避免的，但有三种降低误差的办法，即寻找代理回应、调整统计资料并调查无回应者。使用非概率样本也应注意消除偏误。

资料收集方式的选择，诸如选择问卷邮寄、网络、电话、个人访谈还是群体填答，与抽样、研究主题、样本自身的特点以及可调配的人员和设备都直接相关；同回应率、问题形式和调查成本也有关系。每种资料收集方式各有其优缺点。尽管大多数调查只使用一种方法来收集资料，但是，考虑多种资料收集方式的集合使用，在节约总成本或者扬长避短方面可能更有优势。

社会调查的一个特点是问题的答案可以被测量。便于测量的问题需要良好的设计。一般而言，人们并不仅仅对调查所得的回答本身感兴趣，更重要的是这些答案与被测量对象之间的关系如何。设计出的一个好问题能使得到的答案与研究人员所欲测量的内容最大相关。因此，提高答案的信度和效度就非常重要。这方面做的事情主要有：要有恰当的表述并确保应答者对问题的理解相同，因此要检查问题表述是否完整、是否简洁、是否准确，杜绝界定不明的术语以免产生不必要的误解或歧义；要避免复合问题的出现。目前定类、定序、定距和定比这四种测量方法对应四种待分析的资料，从这个角度来看，对问题的期望也能八九不离十。例如，设计的问题是开放式的还

是封闭式的。但这里要谨慎使用“同意—不同意”形式的问题，因为实际的情形可能要复杂得多。效度问题更为复杂。福勒在这里提到了很多改进效度的建议。这些建议得自一个长达20年的评估和提高测量方法的项目。例如，通过比较口述事实和各种文本记录发现应答者很准确地回答了很多问题，那些未准确回应的原因主要是应答者不明白问题的含义、答不上来、不愿意在访谈中回答等。针对这些原因，可以采取一些措施来加以克服：修改问题，确定并对最了解情况的家庭成员进行访谈、给应答者向其他家庭成员咨询的机会、帮助应答者对答案进行评估等。来自其他研究项目的建议也很有价值，如对敏感问题，可以考虑使主观判断最小化、使精确性最大化、使用自填问卷的方式、匿名处理等。对于主观性的问题，提高其测量效度的措施则一直比较稳定，不外乎尽可能安排单一的问答、答案分类较多且严格排序、找出不同样本同一问题的答案之间的联系和区别来回应主观测量的相对性。

在进行具体的社会调查之前有必要对调查问题和调查方式进行评估。这方面的评估主要有确定调查目标、设计并测试问题，可通过核心小组讨论、初步的问题探讨、系统地复查问题以及实验室访谈、安排好调查工具的具体设计、现场前测等方式来实现。福勒在讨论具体的访谈调查后指出在社会调查完成后还要防止在给资料编码和归并的过程中的误差。他详细讨论了访谈调查的实施过程和注意事项。由于访谈员的素质决定了访谈调查的回应率，而且他们还负责引导与激励受访人并以标准化的不偏不倚的方式主持访谈和答问过程，因此对访谈员的征募、选拔、培训和监督就显得尤为重要。另外，还要对访谈过程进行监控，例如，要求访谈员对受访人进行标准化的介绍并进行标准化的访谈；为了检测访谈的真实性，可以给所有的受访人邮寄简短的问卷询问他们对访谈的意见，也可以要求访谈员抄回受访人的电话来检查访谈情况。这些措施并不一定能完全防止访谈员及访谈过程所产生的调查误差，因此从长远的角度来看，建立一支对研究质量和研究机构负责的常规

访谈员队伍至关重要。现在中国内地有很多调查都是教师委托学生在某些时间进行的社会调查，例如，很多调查安排在寒暑假进行，但是很多学生只是社会调查的局外人，因为教师要的只是他提供的数据，没有将这些学生吸纳到具体的研究情境中，也没有适当的激励，因此很多调查只是例行公事。根据这样的调查所得出来的结论当然不能让人信服并引起大家的注意。

中国内地现在还盛行任务型的课题研究，很多研究机构的学者忙于申请课题，一旦课题申请通过，就将其束之高阁，最后随便写点东西交差，当然这同严肃的社会调查与研究是不相容的。在课题申请中所出现的投机倾向是值得警惕的，例如，有的申请人对申请报告研究透彻，知道怎么样能够顺利通过，他们首先在课题设计上尽量同当局的精神挂上钩，然后最关键的是利用社会调查的形式来欺骗自己和别人，强调自己的研究要进行正规的社会调查，并煞有介事地设计了常见的调查方法，整个框架看上去特别符合人们对科学的来自实际调查的知识获得的期许，但是，这种花架子的社会调查如果有专业且资深的调查方法研究专家来评估，则不堪一击。最关键的是，如果运用福勒在《调查研究方法》这本书中提到的种种误差出现的认识和规避办法来检测这样的课题申报，将使一大批伪社会调查研究现出原形。最可笑的是，在中国内地一些高等研究机构的青年研究项目关起门来自己评估，最后还以“国情调研”之类的名义出版堂而皇之地欺世盗名。

这就涉及调查研究中的道德问题。其实，上述中国内地当代社会调查研究中的现象还不仅仅是调查研究中的道德问题，更是社会科学研究中的一般伦理问题。因为调查研究中的道德问题还是比较具体的，它主要涉及对应答者、受访人的积极告知、保护、奖励以及对访谈员的保护和负责精神。因此，最后势必要求研究者在撰写调查报告的时候就自己所运用的调查方法会影响估计值的种种细节做必要的陈述，同时还要求他们演示有关数值的精度和准确度的计算方法。因为这是社会调

查研究的基本义务。然而,这些义务在中国内地相当多的研究报告中都没有呈现。所有这些方法论上的问题,如果从调查误差来源与有效规避的角度进行总体研究设计上的推敲是完全可以洞悉的。

万卷方法总书目

万卷方法是我国第一套系统介绍社会科学研究方法的大型丛书,来自中国社科院、北京大学等研究机构和高校的近两百余名学者参与了丛书的写作和翻译工作。至今已出版图书51个品种,其中绝大多数是2007年以来出版的新书。

2009年版

1 **量化研究与统计分析:SPSS中文视窗版数据分析范例解析**
邱皓政著
书号:978-7-5624-4821-1

2 **调查研究方法(校订新译本)**
福勒著　孙振东等译　沈崇鳞校
定价:25
书号:978-7-5624-3289-0

3 **质性研究中的访谈:教育与社会科学研究指南**
赛德曼著　周海涛译
书号:978-7-5624-4679-8

4 **案例研究方法的应用(校订新译本)**
殷著　周海涛译　定价:25
书号:978-7-5624-3278-4

5 **教育研究的方法论探索**
孙振东著　定价:38
书号:978-7-5624-4649-1

6 **参与观察法**
乔金森著　张小山译　定价:25
书号:978-7-5624-4616-3

7 **分析社会情境:质性观察和分析方法**
洛夫兰德等著　林小英译　定价:37
书号:978-7-5624-4690-3

8 **建构扎根理论:质性研究实践指南**
卡麦兹著　边国英译　陈向明校
定价:33
书号:978-7-5624-4747-4

9 **公共管理定量分析:方法与技术(第2版)**
袁政　编著　定价:33

10 **AMOUS与研究方法**
荣泰生著　定价:33
书号:978-7-5624-4806-8

11 **文化研究:民族志方法与生活文化**
格雷著　许梦云译　高丙中校　定价:33
书号:978-7-5624-4698-0

12 **质性研究方法:健康及相关专业研究指南**
普拉尼等著　郑显兰译　定价:49
书号:978-7-5624-4720-7

13 **如何做质性研究**
希尔弗曼著　卢晖临等译　定价:58
书号:978-7-5624-4697-2

2008年版

14 **社会科学研究的思维要素(第8版)**
赫文著　李涤非译　定价:25
书号:978-7-5624-4465-7

15 **应用Stata做统计分析(Version 9)**
汉密尔顿著　郭志刚译　定价:58
书号:978-7-5624-4483-1

16 **实用抽样方法**
亨利著　沈崇麟译　定价:22
书号:978-7-5624-4487-9

17 **哲学史方法论十四讲**
邓晓芒著　定价:40
书号:978-7-5624-4446-6

18 **质性研究:反思与评论**
陈向明主编　定价:30
书号:978-7-5624-4462-6

19 **社会研究方法**
仇立平著　定价:39
书号:978-7-5624-4456-5

20 **质性资料的分析:方法与实践(第2版)**
米尔斯,休伯曼著　卢晖临译　定价:78
书号:978-7-5624-4426-8

21 **实用数据再分析法(第2版)**
利普西著　刘军译　定价:37
书号:978-7-5624-4296-7

22 **质性研究的伦理**
丁三东译　定价:32
书号:978-7-5624-4304-9

23 **叙事研究:阅读、倾听与理解**
利布里奇著　王红艳译:定价:27
书号:978-7-5624-4303-2

24 **质化方法在教育研究中的应用(第2版)**
麦瑞尔姆著　于泽元译　定价:33
书号:978-7-5624-4349-0

25 爱上统计学(第2版)
萨尔金德著　史玲玲译　定价:45
书号:978-7-5624-4196-0

26 复杂调查设计与分析的实用方法(第2版)
雷同能著　王天夫译　定价:45
书号:978-7-5624-4290-5

27 美国心理协会写作手册(APA格式)(第5版)
美国心理协会著　定价:49
书号:978-7-5624-4130-4

2007年版

28 做自然主义研究:方法指南
欧兰德森著　李涤非译　定价:32
书号:978-7-5624-4259-2

29 多层次模型分析导论(第2版)
Ita kreft 著　郭志刚译　定价:35
书号:978-7-5624-4060-4

30 评估:方法与技术(第7版)
罗希著　邱泽奇译　定价:49
书号:978-7-5624-3994-3

31 焦点团体:应用研究实践指南(第3版)
克鲁杰著　林小英译　定价:29
书号:978-7-5624-3990-5

32 质的研究的设计:一种互动的取向(第2版)
马克斯威尔著　朱光明译,陈向明校
定价:25
书号:978-7-5624-3971-4

33 组织诊断:方法、模型和过程(第3版)
哈里森著　张小山译　定价:25
书号:978-7-5624-3055-1

34 民族志:步步深入(第2版)
费特曼著　龚建华译　定价:25
书号:978-7-5624-3996-7

35 分组比较的统计分析(第2版)
廖福挺著　高勇译　沈崇麟校　定价:35
书号:978-7-5624-3942-4

36 抽样调查设计导论(第2版)
扎加,布莱尔著　沈崇麟译　定价:39
书号:978-7-5624-3943-1

37 定性研究(第1卷):方法论基础(第2版)
邓津等主编　风笑天等译　定价:58
书号:978-7-5624-3851-9

38 定性研究(第2卷):策略与艺术(第2版)
邓津等主编　风笑天等译　定价:48
书号:978-7-5624-3286-9

39 定性研究(第4卷):解释、评估与描述的艺术及定性研究的未来(第2版)
邓津等主编　风笑天等译　定价:38
书号:978-7-5624-3948-6

40 定性研究(第3卷):经验资料收集与分析的方法(第2版)
邓津等主编　风笑天等译　定价:45
书号:978-7-5624-3944-8

41 研究设计与写作指导:定性、定量与混合研究的路径(第2版)
克雷斯威尔著　崔延强译　定价:29
书号:978-7-5624-3644-7

42 社会网络分析法(第2版)
约翰·斯科特著　刘军译　定价:28
书号:978-7-5624-2147-4

43 公共政策内容分析方法:理论与应用
李钢著　定价:25
书号:978-7-5624-3850-2

2007年以前版

44 论教育科学:基于文化哲学的批判与建构
申仁洪著　定价:25
书号:978-7-5624-3641-6

45 复杂性科学的方法论研究
黄欣荣著　定价:25
书号:978-7-5624-3825-0

46 社会科学研究:方法评论
陈向明著　定价:25
书号:978-7-5624-3689-4

47 电话调查方法:抽样、筛选与监控(第2版)
拉弗拉卡斯著　沈崇麟译　定价:15
书号:7-5624-3441-7

48 科学决策方法:从社会科学研究到政策分析
沃恩著　沈崇麟译　定价:15
书号:7-5624-3669-X

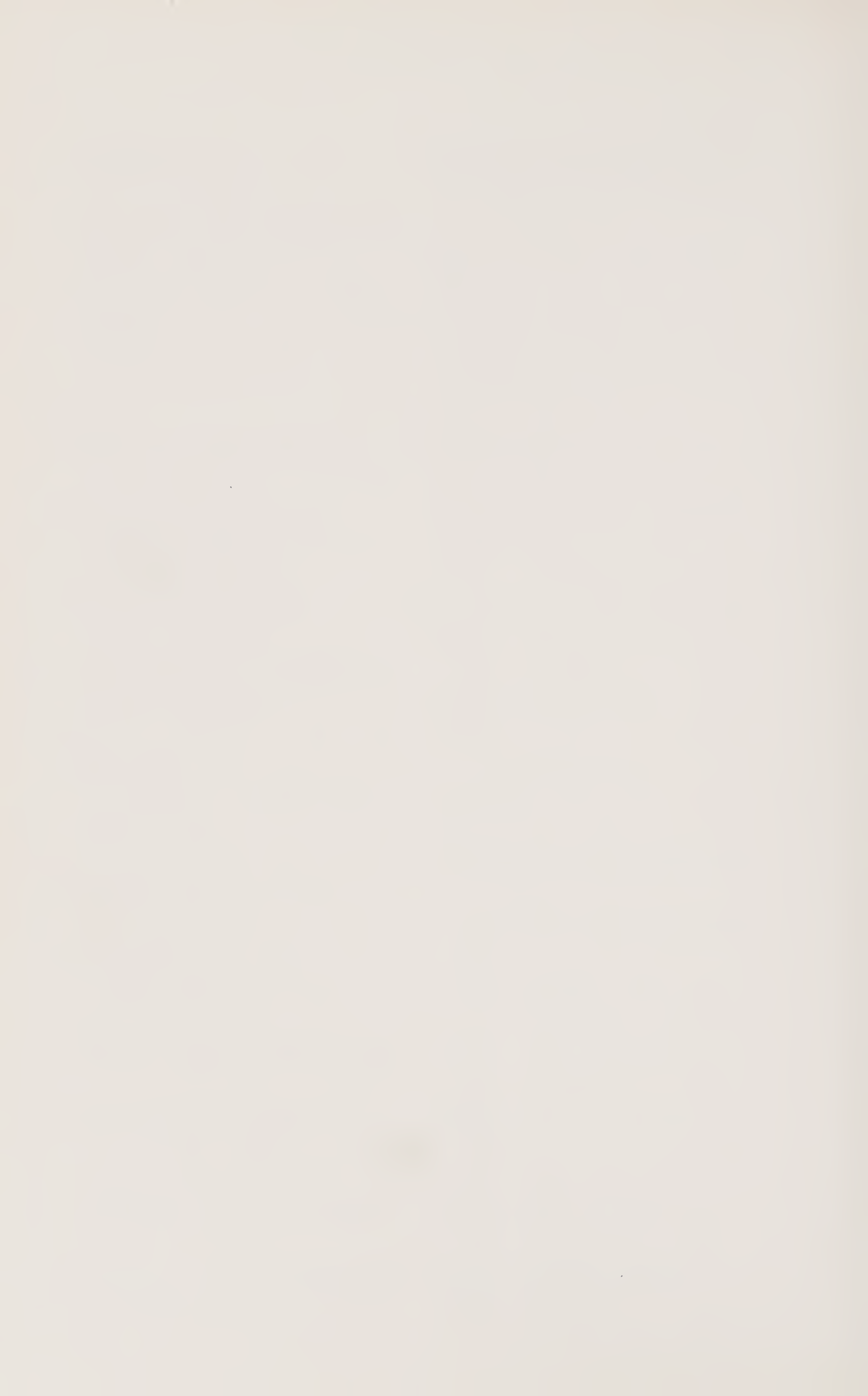
49 研究设计与社会测量导引(第6版)
米勒著　风笑天译　定价:68
书号:978-7-5624-3295-1

50 量表编制:理论与应用(第2版)
德维利斯著　李红等译　定价:15
书号:7-5624-3280-5

51 案例研究:设计与方法(第3版)
殷著　周海涛译　定价:15
书号:978-7-5624-3266-1

52 解释性交往行动主义(第2版)
邓金著　周勇译:定价:15
书号:7-5624-3185-X



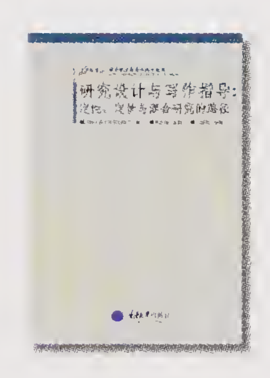


总策划：崔　祝　雷少波



质性研究：反思与评论（第一卷）

这本集中讨论质性研究理论及应用问题的作品内容涉及：对“质性研究”基本概念和理论问题的探讨，以澄清其在社会科学研究中的位置（陈向明）；以研究实例作为分析对象，从方法论角度说明质性方法在研究中的应用（范梅南、康纳利等）；对质性的方法应用、教学等进行反思（林小英等）。本主题将不定期地连续组织出版。



研究设计与写作指导：
定性、定量与混合研究的路径（第2版）

这是一本在西方社会科学界获得广泛赞誉的书，其系统而缜密的研究设计思考、平实易懂的语言、丰富的例子，已经帮助数以十万计社会科学的研究生，顺利完成了他们学位论文的整体设计和写作。



定性研究（全4卷）

本书共4卷，约160万字，内容涉及定性研究的“方法论基础”“策略与艺术”“经验资料收集与分析的方法”“解释、评估与描述的艺术及定性研究的未来”等4个部分。这部在社会科学定性研究方法方面具有里程碑意义的作品，由近80名在国际上享有盛誉的社会科学家，对定性研究的近40个专题进行了讨论，应是所有定性研究方法研习和使用者的案头工具书。



质性资料的分析：方法与实践（第2版）

本书回应了质性研究研究者所面临的迫切需求，即怎样才能从质性数据中引出有效的意义？如何驾驭分析方法？本书诞生二十余年，因其对质性资料分析方法系统、全面、可操作化的论述，而成为西方社会科学质性研究者案头必备的工具书。

本书将在以下方面为您正在或即将开展的调查工作提供帮助：

- 尽可能提高调查的回应率。
- 尽可能提高您在设计题项上的信度和效度。
- 如果您需要使用访谈员协助调查的话，那么本书关于访淡员工作的有关章节必将对您的调查工作有所助益。
- 本书还在调查资料的整理、调查的评价等方面提供可行性建议。

本书作者撰写的《标准化调查访谈》和《改进调查问题》两书，也将由重庆大学出版社出版。这三本书构成了一个体系，是调查研究方法领域最重要的读物之一。

发表及参阅相关讨论，请登录：

万卷方法与学术规范博客圈（http//q.blog.sina.com.cn/fafang）

ISBN 978-7-5624-3289-0



定价：25.00元